The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
System level dynamics of post-translational modifications

Aaron S. Gajadhar\(^{a,b}\) and Forest M. White\(^{a,b,‡}\)

\(^{a}\)Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

\(^{b}\)David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

Attempts to characterize cellular behaviors with static, univariate measurements cannot fully capture biological complexity and lead to an inadequate interpretation of cellular processes. Significant biological insight can be gleaned by considering the contribution of dynamic protein post-translational modifications (PTMs) utilizing systems-level quantitative analysis. High-resolution mass spectrometry coupled with computational modeling of dynamic signal-response relationships is a powerful tool to reveal PTM-mediated regulatory networks. Recent advances using this approach have defined network kinetics of growth factor signaling pathways, identified systems level responses to cytotoxic perturbations, elucidated kinase-substrate relationships, and unraveled the dynamics of PTM cross-talk. Innovations in multiplex measurement capacity, PTM annotation accuracy, and computational integration of datasets promises enhanced resolution of dynamic PTM networks and further insight into biological intricacies.

Introduction

Systems biology aims to identify emergent properties: behaviors, such as cell phenotype, defined by the interaction of many components in the network that are not predictable from the analysis of any single component. The classic dogma of molecular biology, in which information flows from DNA to RNA to proteins to coordinate the development and
function of a cell, does not adequately explain these emergent properties, and also fails to account for the rapid response of biological systems to altered intra- or extra-cellular conditions. A variety of epigenetic phenomena, including protein-protein interactions, chromatin alterations, non-coding RNAs, and post-translational modifications (PTMs), among others, have been implicated in governing cellular responses and phenotypes. Importantly, none of these epigenetic regulatory events can be accurately inferred from genomic information alone [1–4]. The importance of these additional layers of non-genomic regulation cannot be understated; dynamic epigenetic regulatory networks must be considered to fully appreciate the complete nature of a biological system.

One well-studied mechanism by which cells acutely respond and coordinate activities is through post-translational modification (PTM) based cellular signaling networks. The assortment of PTMs in a eukaryotic cell is staggering, with over 600 different protein modification types annotated in the RESID database (September 2013 release; URL http://pir.georgetown.edu/resid/). Protein PTMs can rapidly modulate complex formation, stability, activity, and spatial localization[4]. Integration of this vast array of PTMs ultimately governs cellular information processing and the corresponding cellular behaviors such as migration, apoptosis, and proliferation that are elicited [5]. Of the routinely studied PTMs, phosphorylation, arguably the most abundant PTM in eukaryotic cells, has been shown to drive signal transduction cascades connecting cell surface receptors to resultant cell phenotypes [6–8]. As such, throughout this review special attention will be paid to the role of dynamic phosphorylation in coordinating information flow within the cell and regulating cellular response to cellular perturbations.

Identification of the altered networks underlying emergent properties typically requires systems-level analysis entailing the collection of multivariate data which can yield hypotheses and predictions that are beyond a scientist's intuition [9,10]. Systems level analyses can also clarify paradoxical findings. For example, the dissection of signaling pathways by traditional reductionist approaches can lead to apparent contradictions in the activity or role of individual proteins. The literature is filled with examples of seemingly conflicting results for a number of highly studied signaling components such as Notch and MAPK [11,12]. One landmark paper from the Yaffe group demonstrated the power of systems-level analysis by considering apoptotic regulation by c-Jun N-terminal kinase (JNK) [13]. Through a series of multivariate sampling measurements and data-driven computational modeling, they concluded that phospho-JNK levels could appear anti-apoptotic, proapoptotic, or uninvolved in apoptosis depending on the `signaling state' of the cell. Therefore, a signaling component can have a multivariate nature and response that can only be fully understood by taking into account the context of the network in which it is functioning. Thus, to fully comprehend the multidirectional intracellular interconnections that exist, techniques must be employed to comprehensively detect and quantify multiple components of a network.

The analytical tools available to unravel these details are expanding but ultimately must satisfy minimal requirements to accurately identify and quantify many components within a single analysis, thereby enabling quantitative analysis at a systems level. Mass spectrometry (MS) based proteomic tools fulfill these requirements and have become a mainstay
Dynamics of phosphorylation

Mass spectrometric analysis of protein phosphorylation is an immense analytical challenge. Thousands of proteins are expressed in each cell, with most proteins present in multiple different isoforms, and with each isoform potentially modified at multiple phosphorylation sites [19,20]. Due to the transient nature and varied stoichiometry of each phosphorylation site, it has been very difficult to generate a comprehensive catalog of all protein phosphorylation sites in any given biological sample. It is worth mentioning that, in our opinion, the goal of using systems-level tools to probe phosphorylation must move beyond a
superficial ‘cataloging’ of phosphorylation sites and phosphorylated proteins to understanding and determining the putative phosphorylation marks that are responsive and dynamic in nature. Thoughtful design and execution of experimental studies to understand phospho-dynamics can provide rich insights into the underlying mechanics of a functioning network, reveal unanticipated biological features and profoundly guide future ventures in a way that simple indexing cannot. Here we will highlight a few recent studies have derived biological insight to cellular signaling mechanisms through MS-based network analysis. For instance, a recent report from the Ferrara group utilized a phospho-proteomic strategy to dissect vascular endothelial growth factor (VEGF)-regulated phosphorylation dynamics in a human endothelial cell model [21]. VEGF receptor signaling had been previously shown to trigger the PI3KAKT, Raf-MEK, and Src-FAK pathways. Using a selective enrichment schema with five immunoaffinity motif antibodies combined with quantitative mass tagging, they selectively investigated temporal phosphorylation profiles of these three pathways upon VEGF stimulation. This kinetic sampling allowed identification of a discrete sequence of phosphorylation dynamics revealing membrane-proximal events, followed by nuclear and transcription factor phosphorylation, and terminal events implying establishment of an integrated signaling network. Furthermore, the acquisition of this network-level phosphorylation dataset allowed resolution of molecular conduits by which VEGF signaling may lead to RTK reprogramming in endothelial cells upon pharmacological intervention. Another recent report used MS-based phospho-proteomics, combined with stable isotope labeling, to survey the network changes to protein phosphorylation during cell-cycle recovery from DNA-damaging agents [22]. Their analysis revealed 154 proteins that were quantitatively dynamic during recovery from DNA damage–induced G2 arrest across multiple time points. Interestingly, 84 of these proteins were previously identified in a screen of targets phosphorylated upon DNA-damage, but only two phosphosites were shared between the induction of DNA repair and the recovery from DNA repair. This result suggests that PTMs governing DNA repair machinery are astonishing complex with multisite phosphorylation often regulating diametrically opposed functions on the same protein.

Each of the above examples used mass spectrometry to query changes in signaling networks at the systems level. In many cases, selected kinases within the network have previously been implicated in regulating a given cellular response. To utilize this \textit{a priori} knowledge, analog-sensitive kinases (AS-kinases) and bio-orthogonal ATP analogs can be combined with MS-based phospho-proteomics to identify and quantify direct substrates of given kinases [23,24]. To date, quantification of substrate phosphorylation dynamics has been challenging due to the inability of the ATP analog to penetrate the cell membrane. Semi-permeabilizing the cell membrane allows for addition of the ATP analog to the cell, but can significantly alter cell signaling networks and response to stimulation. Continued development of new methods (e.g. microinjection platforms) should enable more high throughput analysis of kinase substrates across different time points and conditions, thereby providing a much more comprehensive map of the signaling networks. This information would facilitate our understanding of how signaling networks re-wire in the context of different disease states, potentially enabling more directed therapeutic strategies to reset the network.
Dynamics of other PTMs and PTM crosstalk

Deciphering the regulation of other protein PTMs has proven immensely challenging due to a combination of poor affinity capture reagents and the low-abundance/low stoichiometry of many PTMs. Despite these challenges, several studies have now documented proteome-wide lysine acetylation, yet none of these studies have quantified acetylation temporal dynamics [25–27]. Novel affinity capture approaches and advancements in MS technologies have allowed network-wide interrogation of several other PTMs. For instance, protein palmitoylation on cysteine residues has recently been analyzed using bioorthogonal labeling with palmitic acid analog 17-octadecynoic acid (17-ODYA) into the endogenous sites of protein palmitoylation, and quantified with isotopic labeling strategies. This novel approach utilized biotin-azide click chemistry and avidin enrichment for LC-MS analysis [28]. Highly dynamic global protein palmitoylation events were defined on proteins implicated in migration, proliferation and cancer, among others. While sites of protein ubiquitination have recently been described using ubiquitin remnant profiling [29], a recent study has leveraged this technique to screen for ubiquitylated sites that are dynamically regulated in response to ultraviolet irradiation. This study uncovered a vital role for dynamic ubiquitination of PCNA associated factor PAF15 during DNA-damage signaling [30]. Although quantification of ubiquitin remnants provided insight into this system, further improvements to this approach may enable selective profiling of specific monoubiquitination or polyubiquitination linkages to illuminate how these differences regulate system-level signaling. Lastly, a recent study by the Bonaldi group combined heavy methyl isotopic labeling, extensive immunoaffinity enrichment of arginine/lysine methylation, and distinct separation schemes to comprehensively investigate non-histone protein methylation in HeLa cells [31].

Another dimension of PTM regulation applies to the coexistence of multiple modifications on the same proteins and the possibility of functional PTM crosstalk. A combinatorial PTM code has been elegantly deciphered to show how acetylation, phosphorylation, and methylation of histone tails leads to chromatin remodeling and modulation of gene expression [32]. The scope and interplay of co-occurring PTMs on other proteins and pathways is less understood, although several studies have described the interaction of proximal acetylation and phosphorylation sites on selected proteins. Traditionally, probing multiple PTMs simultaneously has been restricted due to compatibility of co-enrichment methods and also by difficulties in detecting peptides concurrently modified with unrelated PTMs. Recently, a novel platform was published to permit i) identification of proteins comodified by ubiquitination and phosphorylation, and ii) identification of proteins where these dual PTM marks were found in close sequence proximity [33]. Using this strategy the extent of ubiquitylation-phosphorylation cross-talk in the context of protein degradation was examined. Intriguingly, spatial constraints appear to be at play for some phosphorylation-ubiquitination modification cross-talk and in some cases there can be a preferential directionality to the regulation of PTMs occurring proximally. Other studies have demonstrated the usefulness of systematic in silico integration of phosphorylation, acetylation, and ubiquitination proteomics datasets to reveal PTM interplay or the importance of PTM conservation as an indication of modifications that are more likely to exhibit cross-talk [34,35].
Novel tools and future challenges

The development of innovative technological advances may enhance our experimental dissection of PTM dynamics on a network scale. For example, the ability to expand the multiplex capacity of MS platform analysis would provide further sampling opportunities and overall depth of quantitative coverage. The recently described neutron encoded chemical and metabolic labeling approaches exploit subtle mass defects that occur due to nuclear binding energy variance of stable isotopes and can extend to 12-plex quantification in the same experiment [36,37].

As the breadth of quantitative data increases, it is worthwhile to call attention to the manner in which this data is annotated and interpreted [38,39]. Although systems level studies serve as a rich community resource, ambiguities in peptide identification or PTM site assignments may exist in these datasets, often to detrimental or misleading effects [40]. Given the intent of using proteomics PTM datasets to derive biological insight, it is imperative that the dynamically regulated PTMs be accurately annotated and sites of modification localized precisely. This task relies on automated database matching and scoring of identifications but can benefit from user-assisted validation strategies [41].

With high quality large-scale databases of dynamic temporal phosphorylation profiles, a large number of computational tools can be applied to the data to imply network structure and gain biological insight. In one interesting approach, the Lauffenburger lab applied a combinatorial bioinformatics algorithm (MCAM) to quantitative tyrosine phosphorylation data describing signaling network response to EGF stimulation [42]. Intriguingly, the most frequently co-clustered phosphorylation sites in this computational approach described site-specific protein-protein interactions, including a novel interaction between EGFR and PDLIM1 [43]. In another approach, the prize-collecting steiner tree algorithm utilizes quantitative phosphorylation data and the protein-protein interactome to infer missing nodes and connect co-regulated phosphorylated proteins into pathways and networks [44]. Additional computational tools to infer kinase-substrate relationships, such as NetworKIN, will also benefit from high quality data describing phosphorylation dynamics [45].

In conclusion, measuring system level dynamics of PTMs has the potential to inform our understanding of everything from fundamental biological processes to decoding the complexity involved in disease networks. Network medicine is an emerging area of interest and has recently witnessed an example of a potential temporal network drug (ie. administration of combination therapy in an order- and time-dependent manner) demonstrating the utility of system level studies [46–48]. Leveraging MS technologies to quantify the dynamic interactions of multiple PTM networks simultaneously will make unique contributions to how systems biology is explored.

References

regulated pathways and functional characterization of network nodes susceptible to pharmacological intervention.


40. White FM. The potential cost of high-throughput proteomics. Sci Signal. 2011; 4:e8. Provocative discussion of the repercussions from low-quality, minimally validated datasets that are available and often serve as a basis for further studies and biological follow-up experiments. Argues for improved data validation strategies and change of emphasis for the objectives of proteomic-based analyses.


Highlights

• Post-translational modifications (PTMs) modulate diverse cellular functions
• PTM signaling networks are dynamic, extensive and interconnect cellular responses
• Mass spectrometry PTM analysis can identify and quantify dynamic network components
• Integration of dynamic PTM signal-response relationships reveals biological insight