Quantitative Dynamics of Phosphoproteome: The Devil Is in the Details

Mogiborahman Salek* and Oreste Acuto

T cell Signaling Laboratory, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom

ABSTRACT: Recent advances in peptide-based (bottom-up) quantitative proteomics and bioinformatics have opened unprecedented opportunities for extensive investigation of cellular proteomes and their dynamics. Here we discuss two approaches currently used to investigate the global dynamics of phosphorylation based on the isolation of phosphorylated proteins or peptides. We evaluate the accuracy of these methodologies to grasp the global dynamics of phosphorylation, and we raise awareness on ambiguities inherent to these analyses. We conclude that further development of targeted approaches should prevent inaccurate conclusions about the nature of biological regulations and in particular kinase-substrate networks.

Mass spectrometry (MS)-based proteomics is a fast-growing research field with constant improvement in sensitivity, enhanced data acquisition speed, and targeted analysis capabilities coupled with ever-refined bioinformatics for automated data analysis.1 This rapid expansion is fostered by the quest to understand the mechanistic bases of biological systems behavior in health and disease.

In addition to deep proteome identification, quantification by MS is crucial in comparative studies (e.g., wild type vs mutant; pharmacological perturbation). MS-based quantification relies on the comparison of the intensity of peptides in a single or separate samples depending on whether biosynthetic2 and chemical labeling3 or label-free quantification4 is employed, respectively. Such an analysis allows one to detect and quantify with sufficient accuracy relative and absolute protein abundance,2,4b dynamics of posttranslational modification (PTM),5 and protein–protein interactions,6 which may accompany changes in the functional state of a cell. The most widely used peptide-based quantitation that allows all the aforementioned quantitative assessment of biological systems is the stable isotope labeling by amino acids in cell culture (SILAC).2,7 In SILAC, differentially labeled peptides are discernible in MS by their mass-to-charge ratio and their intensity that is a measure of the relative peptide abundance in various conditions. For isolated phosphoproteins, the relative protein abundance is estimated by combining quantification of all its detected constituent peptides. In contrast, selectively enriched phosphopeptides quantitation reports abundance changes for individual sites. Both types of analysis allow monitoring global changes in protein phosphorylation from cells in different functional states. The results of such global investigations are usually documented in databases5a,8 that constitute a rich source of information accessible to biologists and biomedical researchers carrying out targeted genetics and biomedical studies on individual proteins or cellular functions and to bioinformaticians and mathematicians to model signaling9 and kinase-substrate networks.10 It is therefore vital to ensure both the accuracy of the data generated and the confidence in their interpretation so that correct conclusions on functional mechanisms can be drawn. Indeed, although sophisticated bioinformatics is deployed to reach the highest confidence of proteomic identification and quantification,11 important ambiguities in data interpretation persists; the observed increase, decrease, or absence of changes in the abundance of peptide phosphorylation need to be interpreted carefully. Here we highlight evident and more recondite ambiguities by discussing how the data from the above-mentioned workflows and the often-combinatorial nature of PTMs may lead to misinterpretations. Further development of peptide-based targeted proteomic analysis12 and its systematic combination with global phosphoproteomic surveys could remove ambiguities and strengthen biological conclusions.

■ QUANTIFYING PROTEIN PHOSPHORYLATION

Protein phosphorylation can be quantified by isolating phosphoproteins or phosphopeptides (Figure 1). Phosphorylated proteins can be effectively enriched by antiphosphotyrosine antibodies (anti-pY Abs) or by protein domains that bind to either phospho-tyrosines (e.g., Src homology 2 (SH2) domain8) or to phospho-serines/threonines (e.g., 14-3-3 proteins;13 (FHA) domain14). Captured phosphoproteins are then subjected to protease digestion (usually trypsin) and liquid chromatography−tandem mass spectrometry (LC−MS/MS) analysis (Figure 1). Changes in phosphoprotein abundance occurring, for instance, during cell stimulation are computed by...
averaging the variation in all detected peptides from a given protein whether modified or unmodified (Figure 1). In contrast, quantitative assessment of phosphorylation changes in selectively enriched peptides entails a supposedly unbiased enrichment of all types of phosphorylation (pS/T/Y), e.g., by TiO2 or IMAC (Figure 1) and usually allows monitoring a larger panel of phosphorylated peptides, hence, a wider palette of cellular functions regulated by phosphorylation. In both workflows, temporal dynamics of phosphorylation can be determined directly from the data for cellular activation of short time spans (0 < t < 30 min), during which the contribution of de novo protein synthesis is negligible. However, for longer time spans the contribution of newly synthetized proteins needs to be monitored in parallel.15

Each analytical workflow has its strengths and weaknesses. Particularly, quantification based on the isolation of phosphoprotein leads to more robust statistics for identification and quantification (based on at least three peptides) while the analysis of selectively enriched phosphopeptides has higher resolution power for precise localization and quantification of the modification at the single amino acid (site) level (Figure 1). Below, we discuss in detail potential limitations and drawbacks of these two approaches one should be aware of and suggest how to address them, by MS-based targeted experiments.

Quantification Based on Phosphoprotein Isolation. Several studies,7,16 including ours,17 have attempted to quantify changes in protein tyrosine phosphorylation during cellular activation in different systems. As shown in Figure 1 (left side), this approach is often based on selective isolation of pY-proteins by anti-pY Ab prior to proteolysis and MS. Because anti-pY immune-isolation can be almost quantitative (e.g., most of pY-proteins are captured17), changes in relative abundance of the isolated pY-proteins at various phases of cellular activation can be assumed to represent quantitative changes in phosphorylation.7,17

On the basis of our own experience, a number of considerations can be made on these procedures. We noticed that mild conditions of cell lysis and fast anti-pY immunoprecipitations isolate not only pY-phosphorylated but also non-pY-phosphorylated proteins.17 The latter may have variant or invariant profiles of activation. In the group with variant profiles, the nonphosphorylated proteins are most often the actual interactors of phosphorylated proteins with closely related profiles. This occurs since large protein ensembles can be formed through intricate connectivity established by modular protein−protein interaction domains and protein adaptors (e.g., SH2, PTB, SH3, WW, BRCT domains or 14-3-3- and GRB2-family proteins). The large excess and relatively high affinity of the anti-pY Ab generally used in these experiments may be expected to dissociate pY-mediated protein complexes during immune-isolation. However, highly cooperative binding through multiple domains (e.g., SH2 and phosphorylation-independent association) may resist interferences by detergent and anti-pY Ab. Isolation of stimulation-induced protein complexes, rather than individual pY-phosphorylated proteins, can be advantageous for information-rich gathering on biological functions of unknown proteins,17 provided that nonspecific co-isolation is excluded by rigorous negative controls and/or by self-evident biological incongruity. Binding after cell lysis of proteins extraneous to such large complexes cannot be formally excluded, though it is much more likely to occur with the matrix (Ab and beads) used for the isolation.18 In most cases, contaminants will be discarded due to their invariant profiles during cell stimulation and/or inconsistent variation in different biological replicas.

Figure 1. Quantitative analysis of phosphorylation. Phosphorylation can be quantified based on the isolation of phosphorylated proteins or peptides. At the protein level (at the left), changes in temporal abundance (0 < t < 30) subsequent to IP using anti-pY Ab are an average of all its identified constituent peptides (identified peptides in red used for quantification can be modified or not). Site-specific resolution could be achieved when analyzing enriched phosphopeptides from SCX fractions using TiO2 (titanium oxide nanobeads) or immobilized metal-ion affinity chromatography (IMAC).
However, it is important to notice that invariant profiles may contain in some cases proteins known by other approaches (e.g., pY immunoblot) to undergo substantial increase in phosphorylation upon cell stimulation. This is illustrated in Figure 2 for the CD3-TCR complex\textsuperscript{17} variation in the abundance of phosphorylated CD3 subunits measured after total pY immune-isolation is smaller than the variation of the selectively enriched tyrosine phosphorylated peptide derived from the TCR/CD3-ζ complex. In this instance, the unexpectedly weak increase in CD3-TCR phosphorylation seen by MS could be attributed to the limited access of antibodies to pY sites. Indeed, after TCR stimulation, tyrosine phosphorylation of the immunoreceptor tyrosine-based activation motifs (ITAM) of the CD3-ζ complex, part of a bulky CD3-TCR multimer, recruits ZAP-70 via the nanomolar-range affinity of its tandem SH2 domains, which hampers the pY residues from complete reaction with anti-pY Abs. Therefore, for such complexes, immuno-isolation would not be quantitative, and the increase in the phosphorylation of individual components could not be accurately appreciated by quantitative MS.

Thus, while very abundant information concerning phosphorylation changes in many proteins during cellular signaling can be gathered by quantifying isolated pY-proteins, caution should be exercised in the interpretation of the data. It is important to keep in mind that these data have indicative value directing further corroborative studies.

**Quantification of Peptide/Site-Specific Phosphorylation.** MS-based quantification of phospho-peptides/sites appears to be the method of choice to study global protein phosphorylation dynamics. Indeed, it achieves better structural resolution (quantifies single peptides/sites) and high sensitivity in monitoring a large array of phosphorylations involved in many cellular functions and removes aforementioned ambiguities associated with the analysis based on the isolation of pY-proteins. As shown in Figure 1 (right side), the current workflow consists in reducing the complexity of the total protein digest by strong cation exchange chromatography (SCX) followed by phosphopeptide enrichment using TiO$_2$ (or IMAC) nanobeads prior to LC–MS/MS measurement.

However, in spite of the attractiveness of the possibility to correlate many cellular phosphorylation changes to modulation in cellular functions, intrinsic limitation of this large-scale approach need to be underlined.

To start with, identification and quantification are often based on only a few or, at times, just one peptide(s) so that peptide-specific quantification may suffer from weak statistical power. Some enriched phosphopeptides might not be detectable by LC–MS/MS due to their physicochemical properties (digestibility, size, hydrophobicity, isoelectric point, efficiency of ionization, fragmentation behavior, etc.) thus limiting detection and/or quantification of some important phosphoproteins. We observed this behavior in adapter proteins such as SLP-76 and LAT, both critical in TCR signaling networks, that can be quantified by protein anti-pY IP, while remain refractory to the peptide-specific quantitation. For instance, pY sites in SLP-76 remained undetectable even after using various single or combination of multiple proteases [unstated details in ref 19]. In such cases, protein isolation prior to quantitation remains the method of choice. Importantly, quantitating variations at single sites is not always possible and 30% of detected peptides carry multiple phosphorylations,\textsuperscript{5a} for which measuring contribution of each site to the total variation of phosphopeptide may not be feasible. Given the importance of multiple phosphorylations in biological contexts, analysis and interpretation of such peptides need particular attention.\textsuperscript{12c}

Phosphorylation-dependent cellular signaling is activated by stimuli that perturb kinase-phosphatase equilibrium. However, in MS data an increase in the abundance of phosphorylated peptides/sites does not necessarily imply higher phosphoprotein stoichiometry (phosphorylated fraction of a particular protein) as a consequence of higher kinase activity. This point is illustrated in Figure 3A (for the ease of argumentation only short activation times are considered so that \textit{de novo} protein synthesis can be ignored). An apparent increase in the tyrosine phosphorylation of a peptide (in red) may result from PTK activity on the unphosphorylated peptide (green). However, it could equally result from dephosphorylation (e.g., by a serine phosphatase in the example of Figure 3A) of pre-existing doubly phosphorylated peptide (in blue). Similarly, a decrease in the abundance of a singly phosphorylated peptide (Figure 3B, the pY containing peptide, in red) might equally result from the phosphorylation of an additional site (e.g., by a serine/threonine kinase) or the dephosphorylation of the existing site (e.g., by a tyrosine phosphatase). In both cases, to find out which reaction is actually responsible for the apparent change in peptide abundance (Figure 3, increase in part A, decrease in part B), one should quantify all species of the same peptide (modified and unmodified, as shown in Figure 3C,D) by a targeted approach,\textsuperscript{12b} such as multiple reaction monitoring (MRM). Hypothetically, if reactions R2 and R2’ were true then we would expect to see quantitation profiles for each peptide as depicted in parts C and D of Figure 3, respectively. Such an experiment would clarify that the increase (Figure 3A) and decrease (Figure 3B) were due to phosphatase (Figure 3C) and kinase (Figure 3D) reactions, respectively. Importantly, in such hypothetical cases and contrary to an intuitively apparent conclusion from the examples in Figure 3A,B, the stoichiometry of protein tyrosine phosphorylation has not changed.

We suspect that other PTMs, in addition to protein phosphorylation illustrated in aforementioned hypothetical cases, might also contribute to such biases, for instance, ubiquitination and the consequential proteasomal degradation, which have important consequences for biological conclusions drawn from these investigations. As shown in Figure 3 D,C, an
An accurate picture of signaling activities could be captured only if all possible species of phosphorylated peptides and their unmodified counterparts are simultaneously quantified. Therefore, we think that combining targeted analysis using for instance MRM might prove beneficial following the first global exploratory phase of the study.

**CONCLUSIONS AND OUTLOOK**

Quantitative analysis of the entire protein phosphorylation or its individual phosphopeptides by mass spectrometry allows capture snapshots of the signaling systems at various stimulation times. This allows identifying proteins and/or sites involved in regulating cellular signaling at different times and locations. However, identification of phosphorylation by protein isolation bears some uncertainty as to the actual phosphorylation of the identified proteins and fails to detect some real changes. On the other hand, measuring PTMs dynamics at the single amino acid resolution could be also a challenging task. Both approaches could lead in some cases to mistaken conclusions, and therefore further improvement and development of new and complementary analytical approaches may be necessary.

A key issue in large-scale proteomic analysis is the accuracy and precision of the changes measured. Estimating the accuracy depends on the approach used and the scale of the analysis. For instance MS-based analysis of phospho-form distribution of a protein is more accurate than NMR and immunobloting. However it is challenging to measure the accuracy in large-scale analysis. In contrast, independent of the scale of the analysis, very high precision can be achieved by MS in multiple measurements of the same sample, which decreases when comparing multiple biological replicas, (Duskek, O.; Acuto, O.; Salek, M., to be published elsewhere). Overall, in large-scale analysis a realistic aim is to determine the global trend in relative changes and its consistency over multiple biological replicas.

Inherent ambiguities exist concerning the analysis of protein phosphorylation when using peptide-based (bottom-up) proteomics; identification and quantification is achieved at the peptide level and then combined into protein-level information. This analytical concept is subject to ambiguities related to multiple PTMs bearing isoforms that could prevent accurate conclusions to be drawn. New developments in large-
scale analysis of intact proteins open new avenues paving the road to solve weaknesses related to peptide-based proteomics.

We argue that weaknesses in the analysis of protein phosphorylation, in particular, and in general for other PTMs, can be alleviated by corroborative MS-based targeted analysis. Considering all possible species of stimulation-responsive phosphopeptides (multiple phosphorylation) would allow one to comprehensively portray the interplay between kinase and phosphatase networks in cellular signaling systems. In this context, without necessarily embarking into large-scale analysis, it would be of high interest to exploit already existing phosphoproteomics data, documented in various databases, which can constitute a starting point to undertake targeted studies. A significant application would be to design standardized MRM assays targeting particular signaling pathways in comparative studies using diseased mice models. Signaling pathways in mammalian cells can also be quantitatively studied by phospho-flow and mass flow cytometry technologies that offer the advantage of single-cell analysis. However, realistically these approaches and other more classical biochemical and genetic methodologies remain restricted to a handful of target proteins for which reagents such as, e.g., good antibodies are available. Instead, targeted phosphoproteomics would allow monitoring hundreds of phosphosites involved in particular pathways, thus decisively contributing to unambiguously identify defects causing the onset of the signaling pathologies.

■ AUTHOR INFORMATION

Corresponding Author
E-mail: mogjibs@gmail.com.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Wellcome Trust Grant No. G076558MA and EU-FP7 “Sybilla” No. 201106 to O.A. We thank Prof. Wolf Dieter Lehmann for his continuous support, constructive discussions, and critical reading of the manuscript. We thank Vincent Geoghegan and Dr. Benjamin Thomas for critical reading of the manuscript and suggestions and Dr. Claudia Brockmeyer for help to quantify dynamics of tyrosine phosphorylation in Jurkat T cells.

■ REFERENCES

(1) (a) Cox, J.; Mann, M. Annu. Rev. Biochem. 2011, 80, 273–299. (b) Walther, T. C.; Mann, M. J. Cell Biol. 2010, 190 (4), 491–500.
(2) Ong, S. E.; Blago, B.; Krachatmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Mol. Cell. Proteomics 2002, 1 (5), 376–386.
(3) (a) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Nat. Biotechnol. 1999, 17 (10), 994–999. (b) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Kainovskii, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Mol. Cell. Proteomics 2004, 3 (12), 1154–1169.
(4) (a) Griffin, N. M.; Yu, J.; Long, F.; O’Hare, S.; Li, Y.; Koziol, J. A.; Schnitzer, J. E. Nat. Biotechnol. 2010, 28 (1), 83–89. (b) Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M. Nat. Biotechnol. 2007, 25 (1), 117–124.
(5) (a) Olsen, J. V.; Blago, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Cell 2006, 127 (3), 635–648. (b) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Science 2009, 325 (5942), 834–840.
(6) (a) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. Nat. Biotechnol. 1999, 17 (10), 1030–1032. (b) Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bradago-Nilsson, E.; Wilm, M.; Seraphin, B. Methods 2001, 24 (3), 218–229. (c) Blago, B.; Krachatmarova, I.; Ong, S. E.; Nielsen, M.; Foster, L. J.; Mann, M. Nat. Biotechnol. 2003, 21 (3), 315–318.
(7) Blago, B.; Ong, S. E.; Krachatmarova, I.; Mann, M. Nat. Biotechnol. 2004, 22 (9), 1139–1145.
(8) (a) Ceol, A.; Chatr Aryamonti, A.; Licata, L.; Peluso, D.; Briganti, L.; Perfetto, L.; Castagnoli, L.; Cesareni, G. Nucleic Acids Res. 2010, 38 (Database issue), D532–9. (b) Daniel, H.; Chica, C.; Via, A.; Gould, C. M.; Jensen, L. J.; Gibson, T. J.; Diella, F. Nucleic Acids Res. 2011, 39 (Database issue), D261–7. (c) Snel, B.; Lehmann, G.; Bork, P.; Huynen, M. A. Nucleic Acids Res. 2000, 28 (18), 3442–3444.
(9) (a) Imamura, H.; Yachie, N.; Saito, R.; Ishihama, Y.; Tomita, M. BMC Bioinf. 2010, 11, 232.
(23) Bendall, S. C.; Simonds, E. F.; Qiu, P.; Amir el, A. D.; Krutzik, P. O.; Finck, R.; Bruggner, R. V.; Melamed, R.; Trejo, A.; Ormatsky, O. I.; Balderas, R. S.; Plevritis, S. K.; Sachs, K.; Pe’er, D.; Tanner, S. D.; Nolan, G. P. *Science* **2011**, *332* (6030), 687–696.