

MILESTONE 21

In pursuit of PTMs

In the 1990s, advances in approaches to mass spectrometry were enabling the nascent proteomics field to grow rapidly (Milestone 20). But these improvements also revealed just how much more work would be required to truly understand how the proteome functions. For this, it was essential to study how proteins interact with one another (Milestone 22) and how they are modified, post-translationally, in the cell.

A single gene in the genome does not 'result' in a single protein entity; thousands of enzymes in the cell modify proteins after translation has occurred. These post-translational modifications (PTMs) range from changes as simple as phosphorylation (the addition of a phosphate group) or the formation of a disulfide bond between two cysteine residues, to those as complex as the conjugation of a chain of ubiquitin proteins to the amino acid lysine, or of a branching series of sugars to an asparagine. Such modifications allow for a great diversification of protein chemistry and, by extension, protein function.

In 1995, hot on the heels of their work describing a method for mass spectrometry-based proteomic analysis (Milestone 20), John Yates and colleagues began to explore whether a similar idea could be applied to interpret spectra from unidentified modified peptides. It was a relatively straightforward task to interpret the collision-induced dissociation (CID; Milestone 13) spectrum of an unmodified peptide—just match it to a theoretical CID spectrum generated from a protein-sequence database. Information about PTMs, however, was not captured in

gene-based sequence databases. The Yates group thus adapted their database search algorithm to account for the increased mass of a peptide containing up to three PTMs. Although the researchers did not apply the method on a large scale in the 1995 report, the technique helped to set the stage for the use of sequence database searching as a means to interpret mass spectra produced by modified proteins.

The lack of information about PTMs in sequence databases was only part of the challenge. Another pressing issue was the low abundance of modified peptides in the proteome, which made them difficult to detect by mass spectrometry without using some form of purification. In 2002, Forest White, Donald Hunt and colleagues developed a chemical approach to enrich for phosphorylated peptides from cell lysate. This method involved digesting proteins into peptides, blocking reactive carboxylate groups by converting them to methyl esters and using immobilized metal-affinity chromatography to capture only those peptides containing phosphate groups. Working in yeast, the researchers identified 383 phosphorylation sites in 216 peptides—the first large-scale analysis of the phosphoproteome. The work served as early inspiration for the potpourri of PTM-enrichment approaches that is available today, ranging from chemical techniques to immunoaffinity methods that rely on PTM-specific antibodies, as was first shown in a landmark paper by Michael Comb and colleagues in 2003, describing an antibody to enrich for tyrosine phosphorylation.

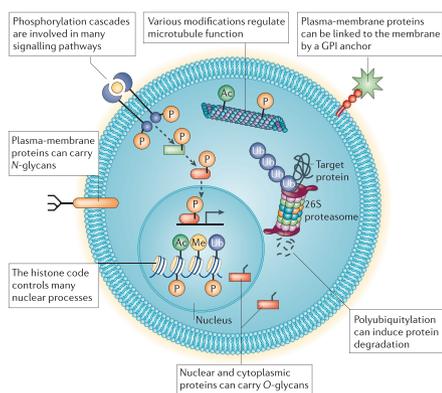
In 2004, Steven Gygi's lab quickly followed in White's and Hunt's footsteps, reporting 2,002 phosphorylation sites in 967 nuclear proteins in a human cell line. Gygi's group achieved this feat by applying strong cation-exchange chromatography to isolate phosphorylated peptides with a charge state of 1⁺ from unmodified peptides with a charge state of 2⁺. They also designed a method to unambiguously identify the phosphorylation site using CID alone: they subjected suspected phosphopeptides to a second fragmentation step to garner additional clues that would enable them to identify the phosphorylated residue. Alternative fragmentation methods (Milestone 13), including electron-capture dissociation, electron-transfer dissociation and

higher-energy C-trap dissociation, would later become important for analyzing protein phosphorylation, as well as other PTMs, on a proteomic scale.

Although some PTMs are irreversible, protein phosphorylation is a transient modification that has an important role in cell-signaling networks. In an early step toward understanding how signaling occurs over time on a proteomic scale, Matthias Mann's group applied a time-course strategy to quantitatively profile the temporal dynamics of 6,600 phosphorylation sites on 2,244 proteins in human cells stimulated with a growth factor. This, along with other work, cemented mass spectrometry as an essential tool in the field of systems biology.

Today, the ability to identify tens of thousands of phosphorylation sites in a single mass spectrometry study (as, for example, the Gygi group showed in a 2010 investigation of the tissue-specific mouse phosphoproteome) is taken almost for granted. Yet it once seemed a near-insurmountable challenge. Quantitative proteomics methods (Milestone 20) have been essential for determining PTM site occupancy and for comparing PTM profiles under different biological conditions. Intrepid scientists have devised a myriad of methods to enrich, detect and profile protein modifications. These include well-known PTMs such as acetylation and ubiquitination, as well as complex carbohydrate modifications (Milestone 12) that play important parts in cell-cell communication but that are extremely challenging to profile in terms of modification site and carbohydrate structure. There are many other modifications whose biological roles are just beginning to be discovered, thanks to the power of mass spectrometry for PTM analysis on a large scale.

Allison Doerr, Senior Editor, Nature Methods



A multitude of post-translational modifications occur in the cell. Image reprinted from Jensen, O.N., *Nat. Rev. Mol. Cell Biol.* **7**, 391-403 (2006).

© 2006, Nature Publishing Group

ORIGINAL RESEARCH PAPERS | Yates, J.R., Eng, J.K., McCormack, A.L. & Schieltz, D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**, 1426-1436 (1995) | Ficarro, S.B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301-305 (2002) | Beausoleil, S.A. *et al.* Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. USA* **101**, 12130-12135 (2004) | Olsen, J.V. *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006)

FURTHER READING | Peng, J. *et al.* A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21**, 921-926 (2003) | Zhang, L. H., Li, X.-J., Martin, D.B. & Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**, 660-666 (2003) | Rush, J. *et al.* Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **23**, 94-101 (2005) | Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840 (2009) | Huttlin, E.L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189 (2010)