Making sure PTMs are not lost after translation

Vivien Marx

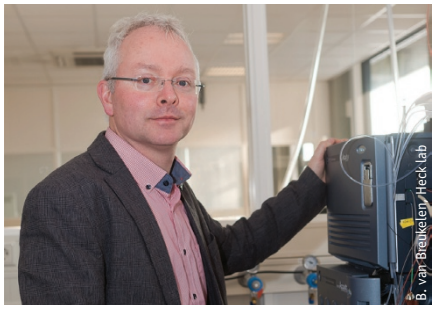
Mass spectrometrists travel top down, middle down and bottom up to study post-translational modifications on proteins and their biological functions.

Proteomics researchers are plotting routes to characterize the cellular function of proteins that have been modified post-translationally and identify how these modifications interact.

To map and quantify modifications on proteins, mass spectrometry is a "central technology in the protein chemist's toolkit," Natalie Ahn, a University of Colorado proteomics researcher and president of the US Human Proteome Organization (HUPO), and her colleagues have pointed out[1]. With various strategies, scientists tune ears and mass spectrometers to track the buzz of proteomic conversation.

Post-translational modifications (PTMs) decorate proteins at many locations along their amino acid chains. Phosphorylation, acetylation and ubiquitylation are among the most prevalent among the over 200 known types of modifications. Quality PTM information is becoming increasingly available in online resources (**Box 1**).

Advancements in mass spectrometry technology and methods are helping scientists discern—on a proteomic scale—how proteins are modified in healthy or diseased cells and are indicating how to characterize the hum of PTM cross-talk. Mass spectrometrist Albert Heck from Utrecht University estimates that there might be half a million phosphorylation sites in a cellular system. A decade ago, researchers could identify around 500 protein modifications in a cell. Nowadays, a large-scale experiment running for a few weeks might yield 30,000 PTMs. "Thirty thousand is impressive, but it's still not half a million," he says.



"The key issue is: we need to zoom in," says Albert Heck.

PTMs can decorate proteins at many locations. Advancements in mass spec technology and methods help scientists discern them on a proteomic scale.

PTMs are usually discovered at so-called substoichiometric abundance, which means they are found on some, but not all, copies of a given protein. And they are often dynamic and transient events. All of this makes them challenging to detect[2].

Despite the considerable technical challenges, an emerging theme in proteomics is the study of how PTMs interact to carry out biological functions. A protein may become phosphorylated, a change that recruits enzymes that ubiquitylate or SUMOylate the protein, marking it for degradation, Heck says.

In jest, protein mass spectrometrist Ole N. Jensen from the University of Southern Denmark and his colleagues call efforts to identify and map PTMs 'modificomics'. Jensen studies the role of acetylation and phosphorylation in diabetes. Mapping phosphorylation has become relatively routine—involving, for example, seeing how five or ten phosphorylation sites are co-regulating a biological function. And these interactions are not just between closest neighbors. "One phosphorylation could be in position 1, and the other could be in position 200, meaning 200 amino acids away," he says.

To pull out relevant biological information, Jensen emphasizes the need for careful experimental design, and for proper controls in particular. A control sample and a test sample with different experimental conditions might each reveal 10,000 phosphorylation sites. "But what you want to find are those 10 or 15 that have biological interest," he says.

Exploring relationships between PTM neighbors near and far is a challenge in the widely used approach called bottom-up proteomics, in which proteins are digested into short peptides prior to mass spectrometry analysis. Scientists hope that the alternate routes they are developing—top-down and middle-down strategies—to query a proteome will be particularly useful for PTM analysis.

Diving in, top down

As Northwestern University mass spectrometrist Neil Kelleher and colleagues

BOX 1 QUALITY DATA DUMP

Proteomics data sets can be shared in many ways, through community-based PTM databases or resources that scientists build on their own. For example, Gygi has a ubiquitylation database (https://gygi.med.harvard.edu/ggbase/).

Many biologists might want to search for just results; others want to mine spectra. "For the tech gurus like me, it is important to see the raw data," says Heck, to allow scientists to evaluate one another's work, for example.

But as scientists plumb existing resources, they also encounter data captured on older instruments with lower resolution, which leads to mixed-quality PTM annotations in databases, says Jensen. Several new initiatives address these issues.

The ProteomeXchange consortium (http://www.proteomexchange.org/) is working on ways to standardize data submission and access to mass spectrometry–based proteomics data in public repositories, says Juan Antonio Vizcaíno, who directs PRIDE (http://www.ebi.ac.uk/pride/), the proteomics data repository at the European Bioinformatics Institute.

"The size of the data sets is currently not the main issue," he says. He and his team are developing tools and pipelines to perform semiautomatic data quality assessment. One of these is PRIDE Inspector (http://code.google.com/p/pride-toolsuite/wiki/PRIDEInspector/, and another is the recently developed PRIDE Cluster (http://www.ebi.ac.uk/pride/cluster/).

In the works at the European Bioinformatics Institute is PRIDE-Q, a resource slated for release later this year that will contain quality-filtered data from PRIDE. This resource addresses issues relating specifically to the validation of data from the quite heterogeneous experiments done over the years, he says.

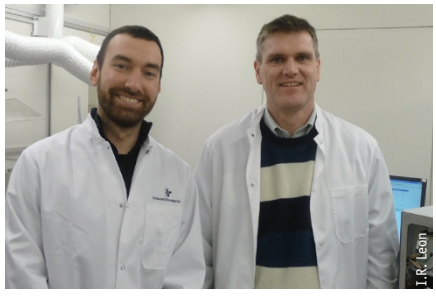
PRIDE also captures and stores PTM data. But the "ambiguity" in the position at which a modification such as phosphorylation has occurred is, for the most part, not taken into account, he says. Repositories are still struggling with ways to capture this information in a standard way.

point out, the bottom-up approach can lead to PTM data that are hard to interpret[3]. But traveling top down to interrogate intact proteins by mass spectrometry has been considered too technically challenging to do on a proteomic scale. However, Kelleher and his team managed to analyze over 3,000 intact proteins, increasing proteome coverage more than 20-fold over previous efforts. They applied a four-dimensional separation method involving two-dimensional liquid electrophoresis and liquid chromatography followed by mass spec analysis.

The increased coverage "provides a path ahead" to interrogate protein structures in human cells and tissues, the authors note. The approach stands to "help lay bare the post-translational logic of intracellular signaling."

Even if it is still the "very, very early days," Jensen shares the excitement among his colleagues about these methods. Current bottom-up methods now robustly map phosphorylation sites. But PTM cross-talk calls for a strategic shift to reach higher proteome coverage, he says. "This is why we are moving toward the middle-down and top-down approaches."

High-performance, high–mass accuracy hybrid tandem mass spectrometers are helping his team annotate PTMs better than even a few years ago, he says. Jensen's team used to apply either quadrupole time-of-flight (Q-TOF) hybrid instruments or low–mass resolution ion traps, using multiple stages of mass analysis in a process called MS/MS for sequencing peptides and localizing their modifications. Advanced Q-TOF and Orbitrap-based systems deliver high speed and perform high-sensitivity sequencing of peptides up to 5 kilodaltons and larger, he says. The right software tools are another crucial factor in PTM data analysis. Jensen's team uses tools developed in other labs, such as Kelleher's ProSight for PTM determination (https://prosightptm.northwestern.edu/) and the search engine Mascot (http://www.matrixscience.com/search_form_select.html), but the scientists are also working on their own software. "We are developing tools to look at cross-talk and PTMs in large peptides," he says.



In jest, Simone Sidoli (left), Ole N. Jensen (right) and colleagues call efforts to identify and map PTMs 'modificomics'.

Zoom in and pull down

Modified proteins are often found in low abundance and can be below the detection limit of current technology. "The key issue is: we need to zoom in," says Heck. Some scientists reach this focus by selecting for study a single protein or individual protein complexes. Another way to focus the PTM hunt is to use targeted antibodies.

As Heck explains, antibodies are emerging that pull down only peptides that are phosphorylated by a particular kinase, which allows researchers to look at phosphorylation changes in a network of 10–20 proteins.

Enriching for low-abundance modified proteins before sending a sample into the mass spec is a step many researchers take. Heck and many others have been working on enrichment strategies for large-scale analysis of particular modifications, such as phosphorylation.

But Heck also notes caveats, such as biases that can occur with some of the existing methods due to physicochemical properties innate to phosphopeptides.

To avoid these biases, he and his team, in collaboration with colleagues at the Chinese Academy of Sciences, have recently used Ti^{4+}-immobilized metal ion affinity chromatography prior to mass spec. "That's our new goodie," he says. The approach lets them monitor several thousand phosphorylation sites in one experiment, "using minute amounts of sample and relatively very little analysis time," he says.

Most recently, the Heck team, along with colleagues in Germany, used this method to perform in-depth proteome analysis of *Trichoplax adhaerens*, one of the first multicellular organisms[4].

They identified many PTMs, including *N*-acetylation, lysine acetylation and phosphorylation. The team believes that their data contribute to ideas about the role of phosphotyrosine signaling in the emergence of multicellularity.

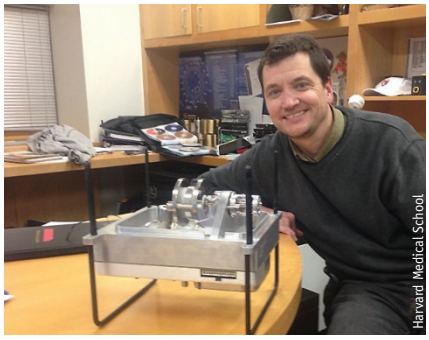
Mass spectrometrist Steven Gygi at Harvard Medical School echoes the call for

careful experimental design when studying PTMs. Among his projects, he focuses on ubiquitylation. As its name suggests, ubiquitin is widely distributed in cells: a broad cross-section of a cell's proteome is subject to ubiquitin modification at some point during the cell's lifetime.

Gygi noticed that digesting proteins into peptides with trypsin revealed a telltale pattern that lit the way to map ubiquitylation on a proteomic scale. The digestion process leaves a "remnant of the ubiquitin molecule," he says. The final two ubiquitin amino acids—glycine followed by glycine—stay covalently attached to the modified amino acid lysine. Given the resulting increase in mass, mass spectrometry can identify peptides with this diglycine remnant[5].

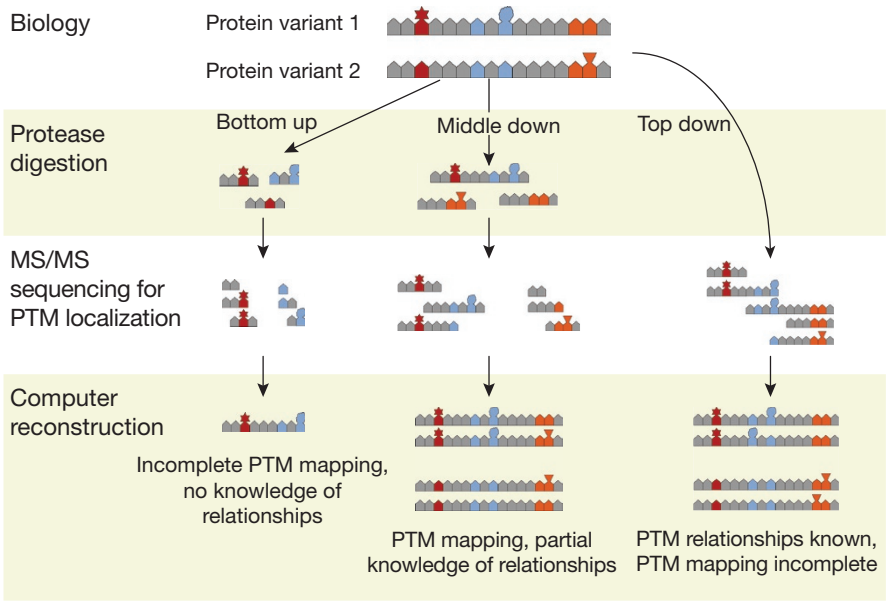
The scientists approached Cell Signaling Technology, which has made an antibody to the diglycine remnant commercially available. Having the antibody in hand represents "a real step forward," Gygi says. Not only does it offer "a way to measure something we never could before," it is helping to track PTM dynamics. "We can now measure global differences in ubiquitin conjugates under different conditions."

In the past, the team has applied quantitative mass spectrometry to chart how these PTM sites change in abundance over different time periods and under changing



"We can now measure global differences in ubiquitin conjugates under different conditions," says Steven Gygi, pictured here in front of a three-dimensional ion trap mass analyzer.

conditions, says biochemist Wade Harper, Gygi's Harvard Medical School colleague. In ongoing work, Harper and his team have been expanding the method to now look at changes in the ubiquitin-modified proteome under varying conditions. One such condition is about a phase of cell metabolism, when the proteasome, a protein complex that breaks down proteins tagged with ubiquitin chains, is inhibited.



Researchers apply different strategies—bottom up, middle down and top down—to localize PTMs. Fragmentation via tandem mass spectrometry (MS/MS) is used to sequence the amino acid chain. In computer-based reconstructions, PTMs are mapped to their locations on the protein.

"This has large implications for understanding ubiquitin biology," Gygi says.

Future PTM world

Overall, the community's ability to identify modified peptides "is in fairly good shape," whereas localizing the modified sites remains challenging, says Ruedi Aebersold from the Swiss Federal Institute of Technology (ETH) in Zurich. Still further from their grasp are questions such as which kinase phosphorylates a modified site, what the biochemical function of a given phosphorylation event is and what characterizes the event's overarching cellular function.

Even if there is a long road ahead, he says his team is achieving "fairly interesting results" in work currently under way as they apply SWATH to the PTM challenge. SWATH is a method to comprehensively query peptide fragment ions within a certain mass range. Aebersold is collaborating with manufacturer AB Sciex to jointly advance both instrumentation and software for the SWATH approach.

As Christie Hunter, AB Sciex director of 'omics applications, explains, her company is focused on quantitation to help scientists measure how modifications change dynamically over time and under varying conditions.

With SWATH, scientists mine for quantitative information in the fragmentation data collected for all peptides falling within a specific mass range. "It's like a digital archive of everything in your sample, or at least everything within the detectable dynamic range of your sample," says Hunter. "We're looking at applying it in the area of post-translational modifications." As new PTM evidence emerges, scientists can reinterrogate the collected fragmentation data at a later point. "We're going to get a whole lot more reuse out of our data than we ever did before," says AB Sciex senior staff scientist Sean Seymour.

Quantifying PTMs from the peptide fragmentation pattern lets scientists "get a tighter look" at where a PTM is localized and allows them to resolve two different

SWATH collects comprehensive fragmentation data that can be reinterrogated, almost like a digital archive of a sample, says AB Sciex's Christie Hunter.

forms of a peptide modified by PTMs, Hunter says. The use of intact peptide mass in PTM analysis has shortcomings, as there may be more than one site where the modification can occur, says Seymour. "You need the fragmentation information to localize the modification where this ambiguity exists," he says.

Future PTM researchers will not need to predetermine the fragmentation approach or energy levels, says Andreas Huhmer.

Advanced mass spectrometry fragmentation, such as collision-induced dissociation and electron-transfer dissociation (ETD), has improved the efficiency of obtaining the amino acid sequence of proteins and finding modification sites. Although ETD can be well suited for specific PTMs, AB Sciex's instruments are all based on collision-induced fragmentation, which is "generically useful" for studying peptides, including those with PTMs, says Hunter.

To enable comprehensive mapping of modifications to particular amino acids in a protein, AB Sciex fosters academic collaborations in the area of algorithm development, Hunter says. Researchers will be able to transfer their data from the company's software into open-source tools. When possible, data converters and other tools will be provided to let them do so.

She believes SWATH simplifies running the mass spectrometer's acquisition mode. The approach makes PTM analysis more accessible to labs. "You don't have to be an MS expert to collect the data," she says.

Ease of use is also a focus at Thermo Fisher Scientific, says Andreas Huhmer, who is the company's 'omics marketing director. Mass spectrometer vendors compete to deliver precision measurement to scientists, quantifying proteins down to a mass error of 10 parts per million. Given that adding a single phosphate group to a 100,000-dalton protein changes its mass by only around 80 daltons, that precision is required in PTM research, he says.

The company's Orbitrap offers high resolution at low mass-to-charge (m/z) ratios, but the resolution "falls off as it gets to higher m/z," whereas the time-of-flight analyzer is better at higher m/z, Huhmer says. In his view though, the Orbitrap offers higher resolution overall than time-of-flight analyzers. His team wants to help apply these methods because peptide enrichment in mass spectrometry calls for a buffer that must become volatile. "If it doesn't, it will just clog up the instrument and you will have a million-dollar instrument just sitting there doing nothing," Huhmer says.

The fragmentation approach used to characterize PTMs also matters. For example, when the standard collision-induced dissociation fragmentation approach is applied to a peptide modified by a phosphate group, the phosphate group pops off. "You have no idea where that thing came from because the peptide itself hasn't fragmented," he says. A researcher will not know which amino acid has been modified.

Huhmer says Thermo Fisher Scientific's instruments are well suited to PTM analysis due to the ETD functionality built into its mass spectrometers since 2006, a method developed in Don Hunt's lab at the University of Virginia. "It was very clear it was a breakthrough in terms of analyzing PTMs, and we made the decision to invest in that and build that," Huhmer says.

With this technique, the peptide fragments in a way that preserves the position of the phosphate group. "It has changed the way people can look at many, many, many of these phosphorylation events," he says, and at other labile modifications. Although other vendors offer ETD, he feels Thermo's implementation is the most practical.

While working toward making its instrument "smart enough to decide what to do when it has a complex sample that has lots of PTMs," his company also cultivates academic collaborations, which are intensifying in the area of algorithm development, he says. A future PTM researcher will not need to predetermine the fragmentation approach or energy levels. "You tell the instrument, 'I am going to do a phosphopeptide,' and the instrument knows, 'Oh, I need to watch out for these things, and then, when I see them, I need to act in a particular way,'" says Huhmer.

1. Witze, E., Old, W.M., Resing, K.A. & Ahn, N.G. *Nat. Methods* **4**, 798–806 (2007).
2. Altelaar, A.F.M., Munoz, J. & Heck, A.J.R. *Nat. Rev. Genet.* **14**, 35–48 (2013).
3. Tran, J.C. *et al. Nature* **480**, 254–258 (2011).
4. Ringrose, J.H. *Nat. Commun.* **4**, 1408 (2013).
5. Peng, J. & Gygi, S.P. *J. Mass Spectrom.* **36**, 1083–1091 (2001).

Vivien Marx is technology editor for *Nature* and *Nature Methods* (v.marx@us.nature.com).