PROTEOMICS

Absolute proteomics

Two groups provide useful computational tools that pave the way for absolute quantitation in mass spectrometry–based proteomics studies.

A typical mass spectrometry–based proteomics experiment starts out with a crude biological sample containing some number of proteins. The sample is often initially separated to enrich for some protein feature, such as phosphorylation. The proteins in the enriched sample are then digested (typically with trypsin) into peptides. The peptides are further separated by liquid chromatography followed by sequencing with tandem mass spectrometry (MS/MS) and finally are matched to their parent protein through database searching. But what truly happens when peptides are ionized and enter the mass spectrometer is still somewhat of an unsolved mystery. One thing researchers do know, however, is that not all peptides are equally favored under ionization conditions.

This limitation has prevented the study of proteomics from being a truly quantitative science, but it has led to creative developments using stable isotopes for labeling. As there are no real chemical differences between isotopes, say of ^{12}C and ^{13}C, peptides labeled with stable isotopes ionize in exactly the same manner as their otherwise identical counterparts. But one can only draw quantitative conclusions about the expression of the same protein under different conditions; one cannot say in an absolute sense that protein X is more abundant than protein Y.

Because isotopic labeling is also expensive, many researchers are using a label-free method known as spectral counting as a relatively crude (yet cheap) way of quantifying protein abundance, which is calculated using the percentage of peptides observed per protein and the total number of repeat observations of the same peptide. However, "if you just count spectra, you can imagine how that fails in lots of different ways," says Edward Marcotte of the University of Texas, Austin. "Bigger proteins will have more spectra, so you have to correct for the size of the protein. Different peptides are observed with different efficiencies, so proteins that are made of easily observed peptides will be more abundant, artificially, than proteins that are composed of hard-to-observe peptides."

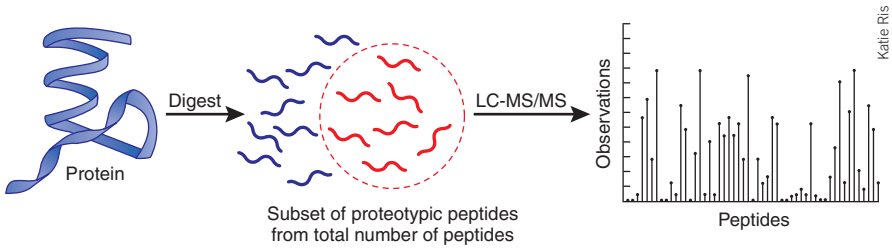

Figure 1 | Illustration of proteotypic peptides. A protein is digested into peptides, but only a subset of proteotypic peptides are observed reproducibly after liquid chromatography (LC)-MS/MS analysis.

To make spectral counting an absolute quantitative tool, "you have to first correct by the probability of observing the peptide," explains Marcotte. He and colleagues therefore developed APEX, a tool to measure absolute protein expression, by correcting for the factors listed above (Lu *et al.*, 2007). They also describe a relative version of APEX, which uses statistics to be able to answer questions about differential protein expression under different conditions. "Anytime you want to look at profiling differences in two samples, it offers an alternative to the isotope labeling methods, and is very fast, cheap and easy," says Marcotte.

Ruedi Aebersold of the Institute of Systems Biology in Seattle, USA and ETH Zurich in Switzerland has also been considering the challenge of absolute quantitation in mass spectrometry–based proteomics. "We noticed that [you always detect] the same peptides from the same protein,…and you don't detect all the peptides you were expecting to see, but it's not random because some always show up and others never show up," he remarks. This observation led the researchers to define these easily ionized peptides as 'proteotypic', or more rigorously, as one that appears in more than 50% of all identifications of the corresponding protein (**Fig. 1**).

Aebersold and colleagues identified more than 16,000 proteotypic peptides from 4,030 yeast proteins, as a result of four large-scale proteomics studies in which more than 600,000 total peptides were identified. They mined the proteotypic peptide sequences for physicochemical information such as charge, hydrophobicity and secondary structure propensity, and used this information to construct a computational tool that can be used to predict proteotypic peptides for any protein, from any organism (Mallick *et al.*, 2007). "[Proteotypic peptides] are detected over and over again," says Aebersold. "We thought we could learn from that and predict which peptides the mass spectrometer should detect from a particular protein and then train the mass spectrometer selectively on those peptides;…then one [begins to] move away from just random sampling [by telling] the mass spectrometer in part what to look for. And then it becomes a much more effective instrument." They found that the computational predictor was about 85% accurate.

Ultimately, Aebersold would like to go beyond predictions by synthesizing a very large number of isotope-labeled proteotypic peptides to use as internal standards in mass spectrometry experiments. This would perhaps be the supreme quantitative resource for the proteomics field.

Allison Doerr

RESEARCH PAPERS

Lu, P. *et al.* Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25**, 117–124 (2007).

Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **25**, 125–131 (2007).