

Mass spectrometry of intact protein complexes

Mass spectrometry technology to detect and characterize large, intact protein complexes is becoming more accessible.

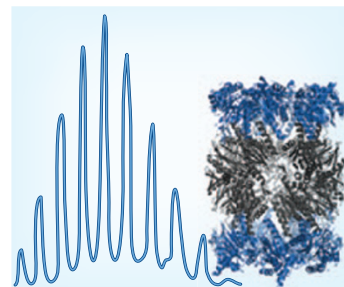
Our Method of the Year 2012 features mass spectrometry as applied to one area, targeted proteome analysis, but this versatile molecular ‘weighing’ technology is well suited to address a variety of scientific questions across disciplines. One interesting area to watch is the application of mass spectrometry to detect and characterize proteins and protein complexes in their intact state.

Most of the time, mass spectrometry analyses of proteins are performed using a ‘shotgun’ strategy: researchers wanting to identify, quantify and characterize proteins in a sample first snip them into peptides, which are more readily detected by the instrument than are intact proteins. Though the shotgun approach has met with wide success for high-throughput proteome profiling, much biological information is lost in the process. Proteins often contain patterns of multiple post-translational modifications that affect

how they function, but when the proteins are chopped up into peptides, these patterns become difficult to interpret. Further, proteins rarely act in isolation; rather, they carry out biological processes by interacting with other proteins and other molecules in the cell. Such information is also lost in a typical shotgun approach.

But by using unique mass spectrometry methods to detect and characterize intact proteins and protein complexes, valuable information providing clues as to a protein’s function is retained. Proteins are gently ionized into the gas phase, which preserves their native three-dimensional structures; under the right conditions, even large, noncovalently bound, soluble and membrane-protein complexes can be observed intact by mass spectrometry.

To date, however, mass spectrometry analysis of intact protein complexes has required highly specialized instrumentation and has therefore been limited to a handful of expert labs. But new developments in 2012 (R.J. Rose *et al.*, *Nat. Methods* **9**, 1084–1086, 2012), extending the use of the popular and powerful Orbitrap mass analyzer to detect large intact protein



The application of mass spectrometry to detect and characterize large, intact protein complexes (such as the yeast 20S proteasome, depicted here) seems likely to grow.

complexes with unprecedented sensitivity, may help bring the technology to a much broader swath of researchers. It will be particularly interesting to watch whether the technology can be merged with advanced ‘top down’ proteomics approaches (J.C. Tran *et al.*, *Nature* **480**, 254–258, 2011)—which seek to profile intact proteins and their post-translational modifications in high throughput—to allow detailed characterization of intact protein complexes on a proteome scale (P.D. Compton and N.L. Kelleher, *Nat. Methods* **9**, 1065–1066, 2012).

Allison Doerr

Machines learn phenotypes

Automated classifiers speed up biological phenotyping.

Detecting variations in the phenotypes of biological samples, be they cells or model organisms, typically involves distinguishing visual patterns. What does the cellular localization of a protein look like upon treatment with signaling modulators? What happens to the morphology of a fly’s head or the locomotion of a mouse when specific genes are mutated or knocked down?

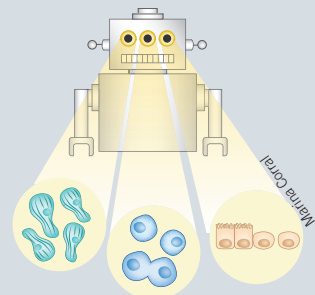
Manually distinguishing such patterns can be tedious for researchers when the number of samples to be examined is very large. But computers can be taught to distinguish biological patterns too.

Machine learning in biology typically involves training an algorithm—or letting it train itself—to automatically classify samples into phenotypic groups. To do this, the classifier must learn which features in the image data are informative for distinguishing the different phenotypic classes. The image data may be static or dynamic, and

they could be of cells or model organisms or any other imaged sample. Phenotypes can be assigned on the basis of prior annotation by an experimenter or can be identified in an unsupervised fashion by the classifier itself. The result is an automated classifier that should accurately distinguish phenotypes of interest in a test sample.

Automated classifiers are particularly apt in the context of large-scale screens, in which tens of thousands of samples, if not more, must be examined. Genome-scale methods to generate mutations or gene knockdowns are readily available. In some systems, comprehensive mutant collections already exist. The bottleneck in large-scale screens lies at the phenotyping stage; it is likely that automated classifiers can help.

In a recent example, Lu and colleagues trained classifiers to automatically identify *Caenorhabditis elegans* with changed synaptic morphology (*Nat. Methods* **9**, 977–980, 2012). By using the classifier in real time as the worms passed through a microfluidic chip, the researchers achieved entirely automated forward genetic screening 100 times faster than would be possible manually.



A computer distinguishes samples with different phenotypes.

Quantitative phenotyping, in which subtle phenotypes must be distinguished or many individuals must be monitored to gather sufficient data for statistical analysis, can also benefit from machine learning. Classifiers can be trained to detect behavioral phenotypes (in model organisms, for instance) and used to automatically map the sequence of these events in many hours of video recordings.

Powerful though it may be, machine learning in biology still faces challenges. But as the tools improve for biologists to implement this approach, it is possible that phenotyping will slowly cease to be the bottleneck in biological analysis.

Natalie de Souza