

Putting the pieces together

Most proteins carry out their function within the cell not as individuals but as components of macromolecular complexes that vary widely in size and complexity. By the mid-1990s, it had become clear that to truly grasp how cells and tissues function would require methods for characterizing how proteins interact with one another. Early efforts by the proteomics community were vital in generating the large datasets of protein-protein interaction (PPI) networks now widely used to connect specific gene products to cellular functions and gene mutations to diseases.

Two earlier developments in mass spectrometry technology formed the basis for the difficult task of systematically identifying individual components of protein complexes: the capacity to establish a peptide's sequence from its mass spectrum and the ability to identify individual peptides within a mixture (**Milestone 20**). Although by the late 1990s proteomics technology had become relatively efficient, better means to systematically isolate native cellular complexes with the required degree of purity and to connect isolated masses to sometimes yet-to-be-characterized gene products were needed.

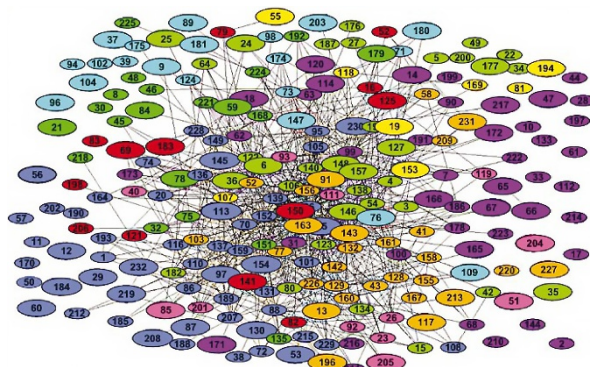
In 1999, John Yates and colleagues demonstrated that large complexes such as the ribosome could be analyzed directly by multidimensional chromatography (see **Milestone 8**) followed by tandem mass spectrometry (see **Milestone 13**), making it possible to identify more than 100 individual components in a single experiment. Still, it was not then clear whether multiprotein assemblies were limited to certain core housekeeping molecular machines, typically around nucleic acids, or were more generally ubiquitous. Although facilitating the identification of components of large complexes, Yates's approach required the ability to purify the target to near homogeneity, and thus was limited to analyzing known complexes.

Around the same time, Bertrand Séraphin and colleagues were developing the tandem affinity purification (TAP) tagging method. This approach, which relied on two N-terminal affinity tags

placed in tandem on the same 'bait' protein, allowed the efficient isolation of proteins directly or indirectly bound as part of the same complex. The use of two affinity tags enabled a two-step purification procedure that substantially limited background contamination while allowing the mild purification conditions needed to maintain the integrity of the protein complex. Séraphin and colleagues demonstrated the method's power by TAP-tagging the yeast protein Snu71p, a component of the U1 small nuclear ribonucleoprotein (snRNP), which they then used to isolate a functional complex containing all the known components of the full U1 snRNP from yeast cells.

To what extent individual components of the proteome act 'socially' as part of macromolecular complexes became the next obvious question. With the ability to isolate native complexes and efficiently identify their components at hand, it was not long before efforts were under way to tackle the protein interactome full on, with the ultimate goal of identifying every protein complex present in a cell. In early 2002, using mass spectrometry, Anne-Claude Gavin, Giulio Superti-Furga and colleagues and Daniel Figeys and colleagues both reported identifying hundreds of protein complexes—many of whose components had no previously defined function—comprising over 25% of the budding yeast proteome. Those analyses provided the first glimpse of the extent to which individual components of the proteome are connected through functional networks. The knowledge of those connections could also be used to infer the function of uncharacterized protein and would eventually pave the way for today's more integrated view of the cellular machinery.

Identifying the components of complexes and the connections between complexes was only the beginning. Those advances were soon followed by demonstrations from groups led by Ruedi Aebersold and Matthias Mann that



Mass spectrometry enables protein interactome analysis. Adapted from Gavin *et al.*, *Nature* **415**, 141-147 (2002).

changes in the abundance of complexes and their subcomponents could also be quantitatively analyzed by mass spectrometry. These advances cemented the idea that proteomics could be used to comprehensively analyze not only how the proteome is organized, but how the cellular machinery functions and responds to perturbations.

Although it has already been over 15 years since these early forays into protein complex identification through proteomics, much remains to be discovered. Mass spectrometry instruments—as well as the methods geared toward efficient isolation and characterization of protein complexes—are still evolving rapidly. Thus it is likely that the field of quantitative interactomics will continue to shape our understanding of normal physiology and disease for some time to come.

Stéphane Larochelle, Associate Editor,
Nature Communications

ORIGINAL RESEARCH PAPERS Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030-1032 (1999) | Gavin, A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147 (2002) | Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180-183 (2002)

FURTHER READING Neubauer, G. *et al.* Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. USA* **94**, 385-390 (1997) | McCormack, A.L. *et al.* Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**, 767-776 (1997) | Link, A.J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676-682 (1999) | Ranish, J.A. *et al.* The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* **33**, 349-355 (2003) | Blagoev, B. *et al.* A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat. Biotechnol.* **21**, 315-318 (2003)