AFFINITY PURIFICATION MASS SPECTROMETRY: PROVIDING STRUCTURAL INSIGHTS INTO THE SOCIAL AND ANTI-SOCIAL BEHAVIOR OF PROTEINS

Keiryn L. Bennett, PhD

Affinity purification is a biochemical approach that was first described by Seraphin *et al. (1)* in 1999 (Figure 1a). Over the last two decades the methodology has been developed further and modified to encompass a diverse range of protein tags. When coupled to mass spectrometry, affinity purification (or AP-MS) enables scientists to extract intact, non-denatured protein complexes from cells and identify the constitutive protein components of these functional complexes. AP-MS has been successfully used to answer specific biological questions on how protein complexes are brought together, how such complexes interact with other protein complexes, and how these complexes may change with perturbations. Disruptions to the normal function of a protein complex may occur as a result of external stimuli, such as a drug or virus. And of course, in the context of disease, AP-MS can provide answers on how protein-protein interactions are altered in response to a specific genetic mutation and what effect this has on the role of the protein. Needless to say, this technology is extremely important in aiding our understanding of not only the stable, undisturbed, healthy human proteome; but how the cellular protein architecture is influenced and adjusted by a specific disease-induced effect. The subsequent architecture of all these proteins within the human proteome can aid us in better understanding the development of disease and ultimately how diseases progress.

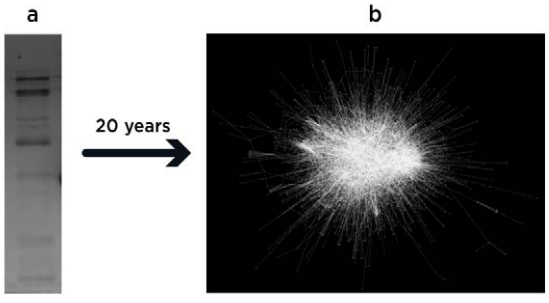

Figure 1. The progress of affinity purification mass spectrometry (AP-MS). In twenty years, AP-MS has evolved from the mapping and discovery of new, individual, functional protein complexes (a) to the mapping and study of entire protein interactomes (b).

A Task of Epic Proportions

A 2017 study led by Ed Huttlin, Steve Gygi and Wade Harper *(2)* from the Harvard Medical School showed how large-scale protein-protein interaction networks generated by AP-MS can be used to study the architecture of the entire human interactome (Figure 1b). The human genome is composed of approximately 20,000 individual protein-coding genes, many of which exist as multiple, alternatively-spliced forms and allelic variants. To create a comprehensive model of protein architecture that reveals how these individual protein assemblies can congregate into functional modules and networks is no small feat. The team achieved this task and compiled the information into a large data repository, termed BioPlex 2.0.,(3) thereby creating a resource that is vitally important for scientists and the general community alike. According to Gygi," Knowing the interactors of a given protein provides [spatial] context for the protein with respect to sub-cellular location within the cell. Further insight is therefore obtained for larger complexes and ultimately complete pathways."

"This approach can provide the missing pieces and alert the scientist to other events that they may not be aware of"
-Steve Gygi

Commenting on AP-MS for structural biologists, Gygi goes on to say that "AP-MS provides an approach where evidence can be obtained for the observation of a direct physical contact [between proteins] including secondary and tertiary binders." In terms of research in structural biology, "This approach can provide the missing pieces and alert the scientist to other events that they may not be aware of, for example, that the phosphorylation of a protein will ultimately affect the structure of the protein under investigation". Similarly, Huttlin says that "New hypotheses can be generated based on associations

amongst [protein] structural domains; moreover, patterns of connectivity can provide structural insight into large protein complexes, like the proteasome, which readily subdivides into its catalytic and regulatory components."

For the non-scientific community, research such as this provides new developments in understanding how a disease initially manifests and ultimately progresses. Huttlin compares the workings of a cell in the context of network biology with the social media interactions of an individual person by saying "With this approach we have essentially created a social network of the cell." A correctly functioning cell is comprised of socially-interacting proteins; whereas in a disease state, several proteins begin to exhibit anti-social behavior.

A Matter of Scale

With a task of such massive proportions, the first goal was to create a reference interactome that placed specific proteins into distinct molecular assemblies. Until now, most genome-wide experimental studies on human protein interactions have relied upon yeast two-hybrid technology or correlation profiling techniques; whilst prior AP-MS-based studies have targeted much narrower areas of the human proteome. Individually and combined, these earlier investigations have mapped only a proportion of the human interactome. Huttlin, Gygi and Harper have taken a tried and tested AP-MS approach to profile protein interactions in the context of a human cell with unprecedented depth and breadth. By targeting an unparalleled number of human proteins for AP-MS analysis using state-of-the-art LC-MS technology, they have been able to map a much larger cross-section of the interactome than had been collectively attempted by other groups and approaches. With this methodology, multiple protein communities have been identified that subsequently enabled them to define and discern several disease networks. To put serious numbers behind the data, from a total of 7,500 AP-MS experiments, Gygi, Huttlin and their fellow colleagues have discovered more than 56,000 interactions that contain more than 29,000 associations that were previously unknown. This information not only provides functional insight into hundreds of poorly-characterized proteins; but also enables prediction of the cellular localization of the proteins.

As with every advancement in scientific research, the approach adopted by Gygi, Huttlin and Harper is not without complications. As Gygi points out, "Dealing with highly-variable levels of bait expression from cell line to cell line can be challenging, as can working with membrane proteins." Similarly, Huttlin states that, "We have also routinely encountered other more technical challenges, including variability of affinity purification over time, intermittent problems with liquid chromatography and mass spectrometry, and so on." Therefore, it was imperative to the success of the project that "A rigid quality control process for both the culturing of the cells and for maintaining and operating the instrumentation was in place."

What Can Be Learnt from the Human Interactome?

This immense network of human protein-protein interactome data will enable other researchers to study protein interactions at a systems level. The data in BioPlex 2.0 *(3)* can lead to the generation of new hypotheses and the discovery of previously-unknown functions of proteins. Indeed, data mining of this network has already led scientists to discover new modalities of under-studied proteins. As pointed out by Huttlin, "Bioplex 2.0 provides a solid foundation to use and integrate other -omic data." In this way, scientists can glean deeper insights into their research by, "Incorporating additional data sets followed by data mining and correlation of the obtained protein networks". "Our future plans" says Gygi, "Are to begin to expand the data set by mapping the human interactome in additional disease-relevant cell lines. In this way, we can start to understand how the dynamic human interactome can alter and adapt as a consequence of specific disease phenotypes."

Bioplex 2.0 is freely-available to the community (http://bioplex.hms.harvard.edu).

References

1. G. Rigaut *et al.*, Nat. Biotechnol. 17, 1030–1032 (1999).
2. E. L. Huttlin *et al.*, Nature. 545, 505-509 (2017).
3. E. L. Huttlin *et al.*, Cell. 162, 425-440 (2015).