Editorial

Protein Complexes and Interaction Networks

In the past century, there has been an increasing realization that in order to understand the function of a protein, one must know which other molecules it interacts with. As detailed in the first article of this special issue ("*History of protein-protein interactions: from egg-white to complex networks*", by Pascal Braun and Anne-Claude Gingras), the turn of the century has seen a tremendous increase in conceptual and technological advances that has propelled the study of protein–protein interactions to the forefront of systems biology. Today, two main types of approaches are used for the identification of protein-protein interactions: biochemical purification of the protein in complex with its interactors (followed by identification of the interactors), and genetic engineering of cellular systems to capture interactions (most often in a pairwise manner) without the need for biochemical purification. In the first category is the now widespread combination of affinity purification with mass spectrometry (AP-MS for short; see below), but also approaches such as LUMIER that involves co-expression of two tagged proteins, one bearing an affinity tag for biochemical capture, and one with a luciferase tag for detection via an enzymatic assay using a luminescent substrate (see second article by Pascal Braun, and also the contribution by Ian Taylor and Jeffrey Wrana at the end of the issue). In the second category is the original yeast two hybrid system, that involves transcription of reporter genes following interaction between the bait and prey proteins fused to different modules of a transcription factor. Since the original two hybrid publication, several variations on the theme have been proposed, including reconstitution of the activity of proteins that enable survival on a selection medium (e.g., split murine dihydrofolate reductase), permit downstream activation of transcription (e.g. the split ubiquitin system), or enable visual detection of the interactions in living cells (e.g. split GFP). While in the split GFP system, the two tested proteins bring together two non-fluorescing halves of a fluorescent protein, fusion of baits and prey to fully active fluorescent of bioluminescent proteins that permit resonance energy transfer permit proximity-based colocalization studies (BRET, bioluminescence resonance energy transfer; FRET, fluorescence resonance energy transfer). These techniques are reviewed here by Mandy Lam and Igor Stagljar, in the context of the interactions involving membrane proteins, which present a notoriously difficult problem in the study of protein–protein interactions.

Anne-Claude Gingras and Alexey Nesvizhskii

A Special Issue on protein–protein interactions could not ignore the great potential and need for getting structural information on the interactions. The article by Bernard Liu, Brett Engelmann and Piers Nash reviews the use of peptide arrays to precisely map the preferred interacting sequences for protein domains. The articles by Suk-Joon Hyung and Brandon Ruotolo, and James Bruce discuss the use of crosslinkers to help mapping interaction topologies, and the application of complex dissociation in the gas phase of the mass spectrometer to decipher the assembly of multisubunit complexes. Though these approaches are fairly new and have been most often used in the context of highly purified protein complexes (for which large amounts of material was available), their refinement holds great promise to provide a structural view of protein complex organization.

In this Special Issue, we have placed a particular emphasis on the assessment of the quality of the protein-protein interactions detected, both in so-called "binary" approaches that aim to score interactions between two proteins, and in AP-MS strategies. For binary approaches, an experimental framework for scoring interactions in high-throughput experiments was proposed a few years ago that involves the use of several orthogonal binary assays: the more often an interaction scores positively across the assays, the higher confidence score it is assigned. These approaches are reviewed by Pascal Braun in the second article of the issue. In the case of AP-MS data, there has also been an increasing realization in the field that filtering of interactions was essential, and various experimental and computational approaches were employed to distinguish true interactions from common contaminants. The article

by Wade Dunham, Michael Mullin and Anne-Claude Gingras provides a general overview of the experimental and computational considerations for designing and interpreting AP-MS experiments. This is further explained in the articles by Marlene Oeffinger, and Jean-Philippe Lambert, Tony Pawson and Anne-Claude Gingras with a special emphasis on RNA-protein interactions and interactions involving chromatin-associated proteins, respectively. A particularly efficient approach to identify true interactors is to perform AP-MS experiments in a quantitative manner: this is reviewed in detail here by Laura Trinkle-Mulcahy, with an accent on the isotope labeling strategy SILAC. We note that quantitative proteomics is also particularly appropriate to analyze interaction dynamics, and this topic was covered in several of the articles.

"the turn of the century has seen a tremendous increase in conceptual and technological advances that has propelled the study of protein–protein interactions to the forefront of systems biology"

Computational methods, tools, and databases for processing and extracting biological information from protein-protein interaction datasets are critical, and this special issue would be incomplete without an in-depth review of various computational challenges. As already mentioned above, the first computational task (beyond the standard protein identification step) in the analysis of AP-MS data is the identification of specific protein interaction partners and elimination of false positive interactions resulting from background contaminants. Existing computational and informatics strategies, and available software tools implementing them, are reviewed here by Alexey Nesvizhskii, with a focus on the analysis of label-free (spectral count or intensity-based) AP-MS data. Scoring protein interactions represents only the first step, however. In a typical experiment, the resulting list of high scoring protein interaction pairs also needs to be visualized, annotated, and computationally interrogated in the specific biological context of the study. This often involves comparing the newly generated experimental protein interaction data with prior knowledge of protein interactions stored in public repositories, as well as submission of new data to those databases. There has been an increased effort by the developers of interaction databases to curate data in a comprehensive and transparent manner, and to develop controlled vocabularies and common data formats enabling a greater degree of data exchange and cooperation between these resources (reviewed here by Sandra Orchard). Another common task involves computational analysis of protein interaction networks with a goal to reconstruct protein complexes. While there exists a large body of computational literature devoted to the problem of protein complex reconstruction, most of those studies focused on the analysis of global (i.e. genome-wide) protein interaction datasets (typically from large scale Tandem Affinity Purification studies in the yeast *S. cerevisiae*). As discussed here by Hyungwon Choi, most of the existing computational methods cannot be effectively applied to AP-MS studies focused on a relatively small number of baits. As this type of dataset is increasingly generated by a variety of research groups and in a diversity of organisms, there is a pressing need for developing computational strategies more suitable for modeling such 'local' (as opposed to genome-wide) datasets.

Representing protein interaction data in a graphical format (network visualization) is a commonly performed task. However, making such visualizations meaningful and biologically informative (and not only visually impressive) is challenging, especially in the case of large datasets. Network visualization approaches and the biological interpretation of these depictions are reviewed here by David Fung, Simone Li, Apurv Goel, Seok-Hee Hong, and Marc Wilkins. Further exploring the subject of network modeling, Sara Mostafavi and Quaid Morris discuss how protein interaction networks can be computationally integrated with other types of networks (genetic interactions, miRNA–mRNA interactions, protein–DNA and protein–RNA interactions, etc.) and other types of data (e.g. gene co-expression or shared subcellular localizations), to enable predictions regarding the function of previously uncharacterized proteins. Protein interaction networks can also be combined with structural information as a way to elucidate the biological properties of such networks. Structural information, coupled with machine learning approaches, can also be used for improved prediction of protein interactions. The role of structural information in the context of protein interaction networks is reviewed by Yogesh Hooda and Philip Kim. Finally, Ian Taylor and Jeffrey Wrana discuss the use of protein interaction networks in understanding and possibly targeting human diseases. They explain that with the help of network graph theoretical approaches, global properties of interaction networks can be analyzed, uncovering relationships between the network organization and disease phenotypes.

The analysis of interaction networks and protein complexes is still a fast developing field. Nevertheless, one important point is becoming increasingly clear: the different experimental approaches mentioned above do provide complementary information, and are all important to generate a clear picture of the interactome. All these approaches, however, also require that proper controls and data filtering are incorporated in each experiment, and that resulting interaction networks are analyzed using robust computational strategies. Rather than fighting over whose approach is best, it may be more productive to ask: what experimental technique should be used to generate (and validate) the interaction data in this particular case, and what computational strategy can be applied to most effectively analyze and interpret these data? We hope that the articles included in this issue will help readers in answering this question for their own experiments.

Anne-Claude Gingras

Alexey Nesvizhskii

