Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Mapping, modeling, and characterization of protein–protein interactions on a proteomic scale

TM Cafarelli[1,2,3,*], A Desbuleux[1,2,3,4,*], Y Wang[1,2,3], SG Choi[1,2,3], D De Ridder[1,2,3] and M Vidal[1,3]

Proteins effect a number of biological functions, from cellular signaling, organization, mobility, and transport to catalyzing biochemical reactions and coordinating an immune response. These varied functions are often dependent upon macromolecular interactions, particularly with other proteins. Small-scale studies in the scientific literature report protein–protein interactions (PPIs), but slowly and with bias towards well-studied proteins. In an era where genomic sequence is readily available, deducing genotype–phenotype relationships requires an understanding of protein connectivity at proteome-scale. A proteome-scale map of the protein–protein interaction network provides a global view of cellular organization and function. Here, we discuss a summary of methods for building proteome-scale interactome maps and the current status and implications of mapping achievements. Not only do interactome maps serve as a reference, detailing global cellular function and organization patterns, but they can also reveal the mechanisms altered by disease alleles, highlight the patterns of interaction rewiring across evolution, and help pinpoint biologically and therapeutically relevant proteins. Despite the considerable strides made in proteome-wide mapping, several technical challenges persist. Therefore, future considerations that impact current mapping efforts are also discussed.

Addresses
[1] Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, USA
[2] Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA USA
[3] Department of Genetics, Harvard Medical School, Boston, MA, USA
[4] GIGA-R, University of Liège, Liège, Belgium

Corresponding author: Cafarelli, TM (TizianaM_Cafarelli@dfci.harvard.edu)
[*] These authors contributed equally.

Current Opinion in Structural Biology 2017, **44**:201–210

This review comes from a themed issue on **Sequences and topology**

Edited by **Ramanathan Sowdhamini** and **Kenji Mizuguchi**

http://dx.doi.org/10.1016/j.sbi.2017.05.003

Introduction

As the most important biological building blocks, proteins mainly carry out their functions by interacting with other biological macromolecules, including DNA, RNA, proteins and small molecules such as lipids and metabolites. Protein–protein interactions, in particular, are incredibly diverse, as they execute a myriad of biological functions. Generating a protein–protein interaction network map at proteome-scale reveals the macromolecular connections that underlie the biology of the cell [1]. Indeed, in order to explore the link between genotype and phenotype and deduce how genetic changes result in disease, an understanding of the cellular network of physical and functional interactions involving proteins is critical [1–4]. As we look to generate the richest and most complete network map possible, we rely on the integration of experimentally derived and computationally predicted interactions. Characterization and application of existing networks has proven useful and highlights the need for expanded network information [5].

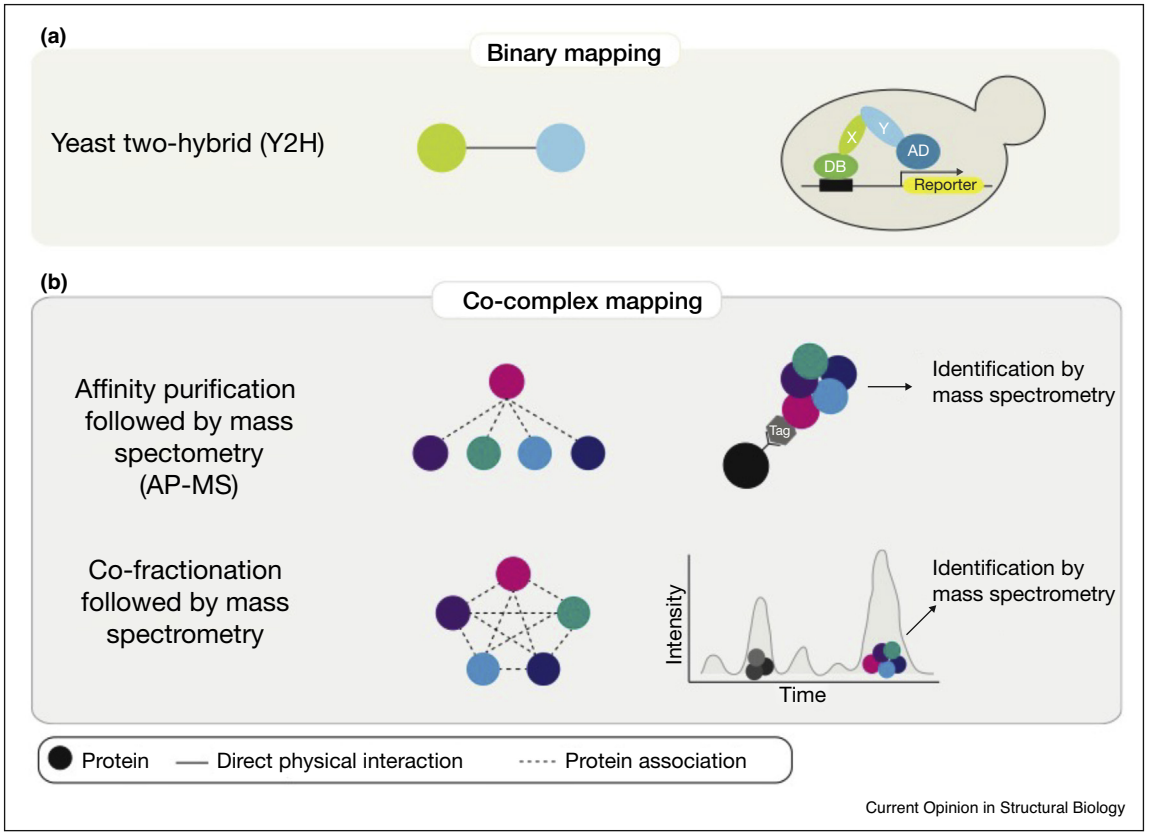
Experimental methods for building a proteome-scale map of the interactome network

There are a number of methods for mapping protein–protein interactions. However, only a few methods are amenable for high-throughput mapping. Recently, proteome-scale interactome maps for human and a number of model organisms have been generated using one of three main techniques. Binary interactome network maps have been generated using yeast two-hybrid (Y2H), and report direct physical interactions. In contrast, co-complex associations can be either direct or indirect protein–protein interactions, and are detected using affinity purification followed by mass spectrometry (AP-MS) or co-fractionation with mass spectrometry (CO-FRAC) (Figure 1). While all of these experimental methods can be adapted to systematically survey the entire proteome, each technique has inherent benefits and limitations.

Binary interaction mapping by yeast two-hybrid (Y2H)

Binary mapping by Y2H detects direct physical interactions between two proteins by the reconstitution of a transcription factor that activates reporter gene expression and promotes yeast cell survival on appropriate selective media (Figure 1a) [6]. Binary interactions identified using Y2H as the primary screening method are validated by a number of orthogonal assays, and data from

Figure 1



Schematic of systematic experimental methods for high-throughput proteome-scale mapping of protein–protein interactions. For each method (left), network representation (middle) and principle (right) are depicted. **(a)** Binary mapping using yeast two-hybrid identifies direct physical interactions between two proteins. **(b)** Co-complex mapping using AP-MS or co-fractionation with mass spectrometry identifies protein associations, which may be either direct or indirect.

such assays indicate that pairs found by this method are of comparable quality as gold standard literature datasets [5,7]. A recent systematic binary mapping study assayed pairs of proteins from a space of ~13 000 × 13 000 human open reading frames (ORFs) and identified ~14 000 protein–protein interactions (PPIs) among ~4300 proteins [8••]. Systematically generated binary maps uniformly identify PPIs in the whole gene space, avoiding sociological bias that may occur in small-scale experiments or literature-curated interactome maps that focus on well-studied genes [8••]. This screening method has therefore proven to be a useful tool, enumerating binary interactions not only for human, but for a number of model organisms as well, including *Saccharomyces cerevisiae* [9–11], *Schizosaccharomyces pombe* [12], *Escherichia coli* [13], *Caenorhabditis elegans* [14,15], and *Arabidopsis thaliana* [16]. While this method is easily scaled and relatively inexpensive, it may fail to capture interactions between proteins which rely on intermediary or scaffold proteins (such as those between protein complex subunits), those involving proteins from specific subcellular compartments (such as membrane proteins), or those which require post-translational modifications [17]. Moreover, this assay requires proteins to be expressed at non-endogenous levels in the yeast nucleus. Such technical requirements may limit the detection of PPIs that require specific protein expression levels (such as protein complexes with strict stoichiometry or stability), or may contribute to the detection of biophysical interactions between proteins that are not endogenously co-expressed or co-localized.

Affinity purification and mass spectrometry (AP-MS)

In interactome mapping by AP-MS, epitope tags are fused to bait proteins, and proteins associated with the bait proteins are purified and identified by mass spectrometry (Figure 1b, top). Two of the latest screening efforts utilizing this method focused on expanding the human interactome network map. The BioPlex dataset reports ~23 700 protein–protein associations (PPAs) among ~7600 proteins, using ~2600 bait proteins overexpressed in HEK293T cells [18••]. An alternative study,