Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Protein–protein interaction networks: the puzzling riches

Shoshana J Wodak[1,2,3], James Vlasblom[1], Andrei L Turinsky[1] and Shuye Pu[1]

While major progress has been achieved in the experimental techniques used for the detection of protein interactions and in the processing and analysis of the vast amount of data that they generate, we still do not understand why the set of identified interactions remains so highly dependent on the particular detection method. Here we present an overview of the major high-throughput experimental methods used to detect interactions and the datasets produced using these methods over the last 10 years. We discuss the challenges of assessing the quality of these datasets, and examine key factors that likely underlie the persistent poor overlap between the interactions detected by different methods. Lastly, we present a brief overview of the literature-curated protein interaction data stored in public databases, which are often relied upon for independent validation of newly derived interaction networks.

Addresses
[1] Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5K 1X8, Canada
[2] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
[3] Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

Corresponding author: Wodak, Shoshana J (Shoshana@sickkids.ca, shoshana.wodak@gmail.com) and

Current Opinion in Structural Biology 2013, **23**:xx–yy

This review comes from a themed issue on **Protein-protein interactions**

Edited by **Joel Janin** and **Alexandre Bonvin**

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2013.08.002

Introduction

Many cellular processes are carried out by physically interacting proteins that often form multi-protein complexes [1]. Characterizing protein interactions is therefore a crucial step in gaining an understanding of how cells function.

Recent technological advances have greatly improved our ability to detect protein interactions [2•,3•]. This has led to an explosive growth of protein–protein interaction (PPI) data derived from many small-scale focused studies, as well as from genome-scale interrogations in organisms such as bacteria, yeast, worm, fly, and human [4–9].

Data from these studies have been used to construct PPI networks, and various properties of these networks have been scrutinized in order to gain biological insights. Topological properties of PPI networks have been analyzed to delineate organizational principles underlying biological systems [10,11]. Other studies have focused on uncovering functional modules [12] and pathways [13,14], or relationships between network connectivity and evolutionary rates [15,16]. Information derived from PPI networks has also been used to infer protein function [17–19] and disease associations [20–22], and to interpret data on gene expression [23] and on single nucleotide polymorphisms (SNP) [24]. The growing interest in PPI networks has also been a powerful incentive to build structural models of the interacting proteome [25•].

This very active exploitation of PPI networks in many areas of research in biology has flourished despite the very abstract nature of these networks, which represent simple connections between proteins and provide no information on the stoichiometry of the underlying interactions, on their temporal or spatial distributions, or on their dependence on post-translational modifications. However, with the proliferation of PPI datasets derived from various experimental techniques, it has been realized that even these simplified descriptions are highly dependent on the particular detection method [26•,27•], raising skepticism that they may not be an authentic enough representation of the functional interaction landscape of the cell (see for example [28••]).

Here we examine some of the reasons fuelling such skepticism. We present an overview of the major high-throughput experimental methods and the PPI datasets produced using these methods over the last 10 years. We discuss the challenge of assessing the quality of these datasets, and scrutinize the technical and biological factors that may underlie the persistent poor overlap between the interactions detected by different methods. To complete the picture of the PPI landscape currently charted by both high-throughput and small-scale experiments, we present a brief overview of the literature-curated protein interaction data stored in public databases.

Genome-scale protein–protein interaction maps derived from experiments

Table 1 lists 16 prominent examples of genome-scale PPI networks derived for major model organisms, of which 8 were published in the last 5 years alone. These networks were built from data collected by two main types of

Please cite this article in press as: Wodak SJ, et al.: Protein–protein interaction networks: the puzzling riches, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.08.002

Table 1

PPI networks derived by high-throughput experimental studies in the last 10 years

Author	Organism	Year	Method	Proteins	Interactions	Complexes
Uetz *et al.* [33]	Viruses *KHSV/VZV*[a]	2006	Y2H	89/69	123/173	NA
Krogan *et al.* [4]	Yeast *S. cerevisiae*	2006	TAP-MS	2708	7123	547
Gavin *et al.* [5]	Yeast *S. cerevisiae*	2006	TAP-MS	1430	6532	491
Collins *et al.* [42••], Pu *et al.* [12]	Yeast *S. cerevisiae*	2007	TAP-MS[b]	1622	9074	400[c]
Yu *et al.* [27•]	Yeast *S. cerevisiae*	2008	Y2H (Union)[d]	2108	2930	NA
Tarassov *et al.* [40]	Yeast *S. cerevisiae*	2008	PCA	1124	2770	NA
Miller *et al.* [38]	Yeast *S. cerevisiae*	2005	Split-ubiquitin	536	1985 (131)	NA
Babu *et al.* [43•]	Yeast *S. cerevisiae*	2012	TAP-MS[e]	2875	13343	720
Simonis *et al.* [87]	Worm *C. elegans*	2009	Y2H (W18)[f]	2528	3864	NA
Guruharsha *et al.* [7]	Fly *Drosophila*	2011	AP-MS	10969	2297	556
Giot *et al.* [32]	Fly *Drosophila*	2003	Y2H	4679	4780	NA
Rual *et al.* [6]	Human	2005	Y2H	1549	2754	NA
Steltzl *et al.* [35]	Human	2005	Y2H	1705	3186	NA
Havugimana *et al.* [8•]	Human	2012	Co-fract.	3006	13 993	622
Hu *et al.* [53]	*E. coli*	2009	AP-MS	1757	5993	443
Arabidopsis Interactome Mapping Consortium [36]	Plant *A. thaliana*	2011	Y2H	2661	5664	NA

The columns (from left to right) list the study and literature reference (col. 1), the organism (col. 2), the year the study was published (col. 3), the detection method used (col. 4), the number of proteins in the network (col. 5), the number of protein interactions (or associations) reported (col. 6), and the number of complexes derived from the works, when available (col. 7). TAP-MS: tandem affinity purification followed by mass spectrometry. AP-MS: affinity purification using a single tag, followed by mass spectrometry. Y2H: yeast two hybrid screens: PCA: protein complementation assays. Split Ubiquitin: membrane yeast two hybrid system. Co-fract: network built by integrating PPI detected by massive co-fractionation and mass spectrometry with various supporting data. See text and Figure 1 for further details on these methods.

[a] The study reports PPI datasets for the Kaposi sarcoma-associated herpes virus (KSHV), and varicella-zoster virus (VZV), respectively. The number of proteins and interactions identified for each of these viruses are listed, separated by a dash.

[b] The reported network was derived by consolidating the raw data from the Gavin *et al.* [5] and Krogan *et al.* [4] studies and applying the Protein Enrichment (PE) scoring scheme to the consolidated data.

[c] The set of 400 complexes was generated by Pu *et al.* [12] from a PPI network with 12035 interactions among 1921 proteins, obtained by thresholding the dataset of Collins *et al.* [42••] with a slightly lower cut off of the Protein Enrichment (PE) score, associated with each link.

[d] This network is a consolidation of the authors' data with the high-quality portions of several earlier Y2H datasets.

[e] This dataset integrates PPI data involving membrane proteins detected in the study, with PPI networks previously derived for yeast soluble proteins. The PE scoring scheme was applied throughout.

[f] The listed number of proteins and PPI for this study is that corresponding to their W18 dataset, which consolidates several earlier datasets.

experimental methods: those that probe binary interactions and those that detect multi-protein complexes.

Salient features of the different methods used to collect the listed datasets are summarized in Figure 1. The detection of binary interactions was carried out using three major techniques: yeast two hybrid (Y2H) screens, split ubiquitin, and protein complementation assays (PCA). The classical Y2H technique involves expressing non-native levels of the tested proteins and detecting their interactions in the yeast nucleus. Y2H was used to derive the first large-scale interaction maps in yeast [29,30] and worm [31], and has since then been applied to other organisms including human [6,32–36]. Split-Ubiquitin [37], a Y2H variant, detects interactions in cellular compartments other than the nucleus and was applied to map out interactions between membrane proteins in yeast [38]. PCA [39] probe the co-localization of two proteins in the cell, and were used to identify thousands of co-localized protein pairs in yeast [40]. Further variants of these methods have been used in small-scale and medium-scale studies of different organisms, including mammals and human (for review see [2•]).

Genome-scale detection of multi-protein complexes uses a fundamentally different approach, involving the coupling of affinity purification methods with mass spectrometry (AP-MS). In AP-MS a tagged protein is expressed in the cell, usually at *in vivo* concentrations, and its associated interaction partners are purified using affinity capture, following which the identity of the purified partners is resolved by mass spectrometry. One of the most powerful methods in this category, which implements a stringent tandem affinity purification protocol referred to as TAP-MS [41] was used to derive several of the more recent large-scale PPI networks for yeast soluble proteins [4,5,42••]. It was recently adapted for the systematic detection of yeast membrane PPI [43•], and extended to medium-scale analyses of PPI in human cells [44]. Less stringent single step purification methods, using simpler tags, have been used to derive recent interaction networks for *E. coli* [34] and the fly [7].

Both AP-MS and binary detection methods probe non-native constructs of the proteins, where tags or larger protein segments are appended to the native polypeptides, potentially altering their properties. Side-stepping this potential problem, a human PPI network was

Please cite this article in press as: Wodak SJ, et al.: Protein–protein interaction networks: the puzzling riches, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.08.002