ELSEVIER

Human protein–protein interaction networks and the value for drug discovery

Heinz Ruffner, Andreas Bauer and Tewis Bouwmeester

Cellzome AG, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

Systematic genome-wide and pathway-specific protein–protein interaction screens have generated a putative, organizing framework of the spatial interconnectivity of a large number of human proteins, including numerous therapeutically relevant disease-associated proteins. The intrinsic value for drug discovery is that these physical protein–protein interaction networks may contribute to a mechanistic understanding of the pathophysiology of disease and can aid in the identification and prioritization of tractable targets and generate hypotheses on how to best drug non-tractable, disease-associated targets. Here, we review the 'therapeutic potential' of the 1st generation sub-genome-scale human interaction networks and disease-associated protein networks generated by yeast two-hybrid screens and affinity purification-mass spectrometry approaches.

Introduction

Traditionally, the spotlight in drug discovery has been on a relatively small number of validated therapeutic target classes, such as G-protein coupled receptors and protein kinases, with well characterized enzymatic and cellular activities that are chemically tractable. The identification and validation of novel therapeutic targets for any disease indication is an inherently difficult, time consuming and expensive process. The biopharmaceutical industry has to date not yet exploited and capitalized on the proteome information encoded by the human genome sequence for novel, innovative therapeutic strategies, which has become available in 2001 by the seminal effort of the human genome project. In the post-genomic era the classic paradigm of tractability and, inherent to that, the scope of drug discovery projects is slowly shifting from a protein-centric view towards a more holistic, pathway-centric view [1]. Ensembles of up to 100 distinct proteins that are physically interconnected and functionally act in concert to transduce extrinsic and intrinsic information are viewed as the target modules for therapeutic intervention. The conceptual advantage of this paradigm shift is that it may not only provide for a molecular taxonomy of disease but may also enable rational, mono- or multi-target therapeutics based on biologicals for extracellular targets and small molecules and RNA interference (RNAi) strategies for intracellular targets. Large-scale, genome-wide protein–protein interaction screens have great potential to expand pathway sub-networks, identify novel therapeutic targets and provide the basis for a molecular understanding of the etiology and progression of disease. In this review, we highlight recent advances in methods analyzing protein interactions, with an emphasis on mass spectrometry-based methods, point out some salient features of the 1st generation human protein interaction networks and use some examples of disease protein networks to illustrate the potential value for drug discovery.

Corresponding author: Bouwmeester, T. (tewis.bouwmeester@cellzome.com)

Methods for the identification of protein–protein interactions

A number of experimental methods, based on distinct, physical principles have been developed to identify protein–protein interactions such as the yeast two-hybrid method (Y2H), affinity purification-mass spectrometry (AP-MS) approaches and protein microarrays [2–5]. The primary modus operandi to identify and map binary, physical protein–protein interactions is the Y2H assay. This method is easy to implement, amenable to automation and relatively cheap. Nearly any protein, including most therapeutically relevant target proteins, expressed as *Bait* protein fused to a DNA-binding domain can be screened against the full complement of a proteome, expressed as individual *Prey* proteins fused to a transcriptional activator domain. The Y2H method has been

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2007.07.011

GLOSSARY

Interactome a protein–protein interaction network that describes the complement of physical interactions detected by a given method.
Genome-scale all proteins encoded by the genome.
Sub-genome-scale a fraction of proteins encoded by the genome.
Hard-wiring constitutive physical interactions that are easy to detect with methods described.
Soft-wiring transient, induced physical interactions, resulting in protein modification that are typically difficult to detect with methods described.
Orthogonal data integration comparison with different high-throughput datasets.
Node protein connected in a network.
Hub protein with many physical interactions (highly connected).
Centrality-lethality rule describes the notion that deletion of a hub protein is more likely to be lethal than deletion of a non-hub protein.

used to generate global protein–protein interaction networks for the bacterial strain *Helicobacter pylori*, the yeast *Saccharomyces cerevisiae*, the fruit fly *Drosophila melanogaster*, the worm *Caenorhabditis elegans* and recently a significant fraction of the human proteome [6–12].

The second major modus operandi, AP-MS is particularly suited to identify and map multi-protein complexes under near to physiological conditions and is therefore complementary to the Y2H method. AP-MS is based on immuno-affinity purification methods, such as the tandem affinity purification (TAP) or other single affinity tags, in conjunction with mass spectrometric protein identification strategies. Epitope-tagged proteins are transduced in immortalized cells and immuno-purified with reagents specific for the tag. All co-purifying, specific ('interactors') and non-specific ('false positives') proteins are identified by mass spectrometric analysis. Alternatively, specific antibodies can be employed to purify endogenous protein complexes from physiological and pathophysiological conditions. The advantage of AP-MS-based approaches is that they provide real-time *in vivo* snapshots of protein assemblies. AP-MS strategies have been applied to generate protein–protein interaction maps for the bacterial strain *Escherichia coli*, the yeast *S. cerevisiae* and recently a small fraction of the human proteome [13–18].

In addition, protein microarrays are being developed to identify binary protein–protein interactions. The concept is based on high-density immobilization of purified, recombinant proteins onto a surface-coated glass slide, which is then probed with fluorescence-labeled target proteins to detect physical interactions. Although of great future potential, a comprehensive proteome array has thus far only been described for *S. cerevisiae* to detect calmodulin interactors [19]. A comparison of the different methods is provided in Table 1.

In general, large-scale datasets suffer from an intrinsically high rate of false positive identifications, necessitating *a posterior* curation by statistical and bioinformatic methods. Various methods, including iterative clustering algorithms have been used to rigorously assess the validity of pair-wise physical interactions [15]. In addition, integration of orthogonal molecular datasets, including data derived from single deficiency and synthetic lethality screens, co-localization and co-expression analyses and co-occurrence of gene ontology terms have been used to further substantiate the validity of individual physical interactions and general network properties (Figure 1).

One major challenge constitutes the experimental validation of specific interactors or complex components as this is mostly a difficult and time-consuming task. The combination of AP-MS with RNAi-mediated knockdown, albeit a low throughput procedure for now, has the potential to validate directly target protein-specific interactions (Figure 2). This method, which has been dubbed quantitative immunoprecipitation combined with knockdown (QUICK) has been successfully used to identify and validate interaction partners of CBL and β-catenin, two signal transduction

TABLE 1

Comparison of different methods used to identify protein–protein interactions

	Yeast two-hybrid method	AP-MS method	Protein microarray method
Key features			
Scaleabilty/throughput	High throughput	Low/medium throughput	Medium/high throughput[a]
Assay set-up	*In vivo* (yeast)	*In vivo* (e.g. human cells)	*In vitro*
Interactions			
Physical interactions	Binary	Binary, protein complexes	Binary
Identification of PTMs	No	Yes	Indirect
Quantitation	No	Yes (e.g. SILAC, iTRAQ)	Yes
Available interactomes (scale)			
Model organisms	Genome-scale	Genome-scale	Genome-scale
	Yeast, *Drosophila*, *C. elegans*	Yeast	Not published yet
Human	Sub-genome-scale	Sub-genome/pathway-scale	Protein-scale
Drug discovery utility			
Applications	Target identification	Target identification	Target identification
	Mechanism of disease	Mechanism of disease	Mechanism of disease
	Protein-compound interaction[b]	Biomarker discovery (e.g. phospho signatures)	Diagnostics[b]
		Protein-compound interaction[b]	Protein-compound interaction[b]

[a] Dependent on the availability of proteome complement of purified proteins (currently only published for yeast) [19].
[b] Applications not covered by this review.