MINIREVIEW

Building Protein-Protein Interaction Networks with Proteomics and Informatics Tools*

Mihaela E. Sardiu[‡] and **Michael P. Washburn**[‡§1]

From the ‡Stowers Institute for Medical Research, Kansas City, Missouri 64110 and the §Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas 66160

The systematic characterization of the whole interactomes of different model organisms has revealed that the eukaryotic proteome is highly interconnected. Therefore, biological research is progressively shifting away from classical approaches that focus only on a few proteins toward whole protein interaction networks to describe the relationship of proteins in biological processes. In this minireview, we survey the most common methods for the systematic identification of protein interactions and exemplify different strategies for the generation of protein interaction networks. In particular, we will focus on the recent development of protein interaction networks derived from quantitative proteomics data sets.

Global efforts in a variety of model organisms have been conducted or are still under way to understand biological processes or whole organisms at the systems level. Among these, a significant focus is on the elucidation of protein interactions because protein interactions are focal points in almost all cellular processes. Within a cell, protein interactions range from pairwise, as in the case of ligand-receptor binding during signal transduction, to the formation of macromolecular protein complexes such as the RNA polymerase complexes or the processosome. However, the vast number of interacting proteins (an average protein interacts with several different partners) and the complexity of protein interactions (for example, many proteins are members of distinct protein complexes) (1–3) require efficient means to display and visualize protein interactions. The main approach used toward this goal is the construction of protein interaction networks. Such networks might either describe local subsets of protein interactions such as the relationship of proteins within a protein complex or under certain biological conditions or eventually depict all global protein interactions on the scale of a whole organism. In this minireview, we will survey the major experimental methods applied for the systematic analysis of protein interactions and examine the alternative approaches to deduce protein interaction networks from the different medium- and large-scale data sets.

Methods for Determination of Protein-Protein Interactions

The yeast two-hybrid system, originally published in 1989 (4), allows the profiling of protein interactions of any two proteins and may be scaled up to cover the whole proteome of an organism. To measure the capability of two proteins to physically interact, one protein (the "bait") is fused to the DNA-binding domain of the yeast protein Gal4, whereas the second protein (the "prey") is fused to the transactivation domain of Gal4. Physical binding of bait and prey is assayed by the activity of a reporter, which is driven by a specific sequence motif recognized by the Gal4 DNA-binding domain (Fig. 1*A*, *panel a*). The yeast two-hybrid system has been successfully applied to screen for the whole protein interactome of different eukaryotic species such as *Saccharomyces cerevisiae* (5, 6), *Caenorhabditis elegans* (7), and *Drosophila melanogaster* (8) as well as for a partial human interactome (9, 10).

The yeast two-hybrid system allows screening for protein interactions of any eukaryotic species under *in vivo* conditions; however, whenever heterologous proteins of higher eukaryotes are assayed, the system will give reliable results only if the post-transcriptional modifications of the two proteins are conserved in yeast. In addition, false results may be observed if additional proteins present in yeast cells positively or negatively affect the binding of the two proteins assayed or if the bait or prey protein itself harbors a transcriptional activator or repressor activity. To overcome this limitation, powerful mammalian two-hybrid systems have been developed (reviewed in Ref. 11). These systems are not as easily scaled as the yeast two-hybrid system and have been used to study more focused networks (reviewed in Ref. 11).

Affinity purification-mass spectrometry (APMS)[2] is based on the biochemical purification of proteins from cell extracts. After purification, the proteins bound (preys) to the protein purified (bait) are determined using MS (Fig. 1*A*, *panel b*). This strategy allows the identification of protein interactions under the physiological conditions within a cell. Several different affinity purification strategies have been employed, the most popular being tandem affinity purification (TAP), in which the bait protein is fused to the IgG-binding domain of *Staphylococcus aureus* protein A and a calmodulin-binding peptide, separated by a tobacco etch virus protease cleavage site (12). TAP-tagged proteins are first purified using IgG beads and, after being released using tobacco etch virus protease, by calmodulin-coated beads. This dual purification strategy strongly minimizes the co-purification of unspecific proteins; however, prey proteins exhibiting a transient or weak association with the bait might be lost. A contributing factor for the popularity of using TAP tags for APMS experiments is the availability of a genome-wide yeast library, in which every endogenous open reading frame is TAP-tagged (13). Using a qualitative APMS strategy, large-scale studies have attempted to define the entire yeast

[1] To whom correspondence should be addressed. E-mail: mpw@stowers.org.

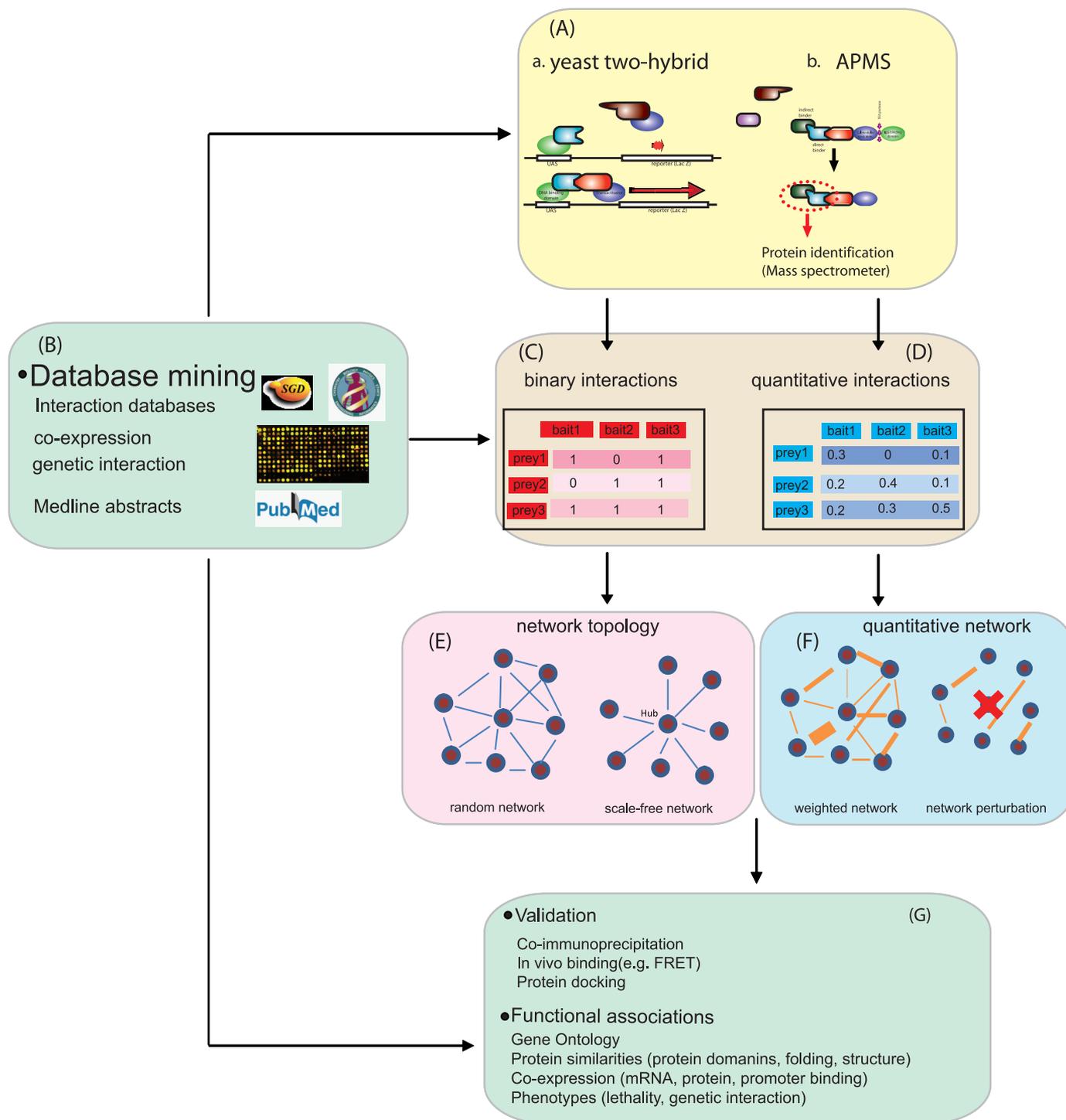[2] The abbreviations used are: APMS, affinity purification-mass spectrometry; TAP, tandem affinity purification.

FIGURE 1. **Protein interaction network analyses.** Protein interaction data can be retrieved either experimentally (*A*) or from database mining (*B*). In database mining, protein interaction databases can use indirect evidence such as functional associations or co-appearance in the same MEDLINE abstract to build networks. The main methods for the experimental identification of protein interactions are the yeast two-hybrid system (*A, panel a*), which delivers direct pairwise interaction data, and affinity purification followed by mass spectrometric identification of the proteins that were bound (either directly or indirectly) to the tagged protein (bait) (*A, panel b*). Binary (obtained from database mining, yeast two-hybrid screens, and APMS) (*C*) or quantitative (quantitative APMS experiments) (*D*) bait-prey matrices are used as input for the construction of protein networks (*E* and *F*). *E*, for protein networks, different topologies are distinguished, *i.e.* random (proteins are evenly connected) or scale-free (only a few proteins, the so-called hubs, harbor the majority of interactions). *F*, quantitative protein interaction networks are typically obtained from quantitative APMS experiments, therefore the intensity of the links represents direct and/or indirect associations. Networks representing more direct interactions between proteins can be obtained by perturbations to the system before performing the quantitative APMS experiments. *G*, the significance of protein interaction networks is typically evaluated 2-fold. Follow-up experiments are performed to verify the interaction obtained from medium- or large-scale screens.

protein interaction network (1, 2). Combining these and other data sets has led to a more detailed map of the yeast protein interaction network (3). However, recent analyses of protein interactions under different physiological conditions within distinct compartments of a cell or upon certain stimuli such as a study of the yeast kinase and phosphatase network have sug-

gested that the definition of the global yeast protein interaction network remains incomplete (14).

In the case of higher vertebrates, for which such libraries of endogenous tagged proteins are not available, protein complexes are generally purified by overexpression of epitope-tagged bait proteins and different epitopes such as HA, FLAG, and Myc (reviewed in Ref. 15). To date, human protein interaction network analyses have been of smaller scale compared with yeast, largely due to time and cost. Examples of human APMS-based protein interaction network experiments are the purification of the TIP49a/b-containing chromatin-remodeling complexes using a combination of HA- and FLAG-fused baits followed by MudPIT (multidimensional protein identification technology) (16) and a study applying a bacterial artificial chromosome-based strategy to overexpress and purify 239 S-peptide-fused baits in mitotic HeLa cells followed by protein identification using SDS gel electrophoresis and MS/MS (17). In addition, the protein interaction networks of the human deubiquitinating enzyme interactome (18) and the human autophagy interactomes have been described (19). Finally, emerging research suggests that antibodies against endogenous proteins may also be used for the analysis of human protein interaction networks (20).

For the identification of weak and transient protein interactions, which in normal purification protocols are lost during the washing steps, chemical cross-linking approaches have been developed and applied to the analysis of the proteasome (21–23). Through reversible cross-linking, even weakly interacting proteins become strongly associated with the bait or the protein complex purified by the formation of covalent bonds and can be biochemically purified. For example, with the use of a His-Bio tag (a peptide containing an *in vivo* biotinylation site flanked by two hexahistidine tags), which remains accessible for purification after formaldehyde-based fixation, this strategy was employed to purify the 26 S proteasome in yeast using Ni^{2+} chelate chromatography and streptavidin-coated beads followed by SDS gel electrophoresis and LC-MS/MS (23). It is likely that as APMS analysis of protein interactions grows, similar strategies will be adopted for the analysis of transient interactions of additional complexes.

Network Evaluation

Using protein interactions determined by any single or a combination of multiple experimental methods and/or data sets, protein interaction networks can be generated, in which each protein is represented by a node, and two proteins are linked with each other if they have been shown to interact. Qualitative data are important for the characterization of network properties such as the density of connection (k), the degree distribution ($P(k)$), the clustering coefficient (C), and the length of the shortest path. These parameters can be used to describe models of biological networks, which under an idealistic approximation can fall into three different classes, *i.e.* random networks (the node degree distribution follows a Poisson distribution, indicating that most nodes have approximately the same number of links) (Fig. 1*E, left*), small-world networks (individual nodes have only a small number of neighbors that can be reached from one another through a few steps), and

scale-free networks (the node degree distribution follows a power law) (Fig. 1*E, right*). The majority of protein interaction networks in yeast, worm, and *Drosophila* seem to be scale-free or follow a truncated power law (24).

To facilitate access to the vast amount of interaction data that can be extracted from the various screens as well as from focused studies, several different groups and consortia have developed very useful public databases such as Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), Biological General Repository for Interaction Datasets (BioGRID), Database of Interacting Proteins (DIP), and Molecular Interactions Database (MINT) and species-specific databases such as *Saccharomyces* Genome Database (SGD) and Human Protein Reference Database (HPRD). Although the protein interactions in the different databases largely overlap, the databases are complementary. For example, STRING not only uses physical protein-protein interaction but also utilizes other evidence such as coexpression or mining of publications. BioGRID distinguishes between low- and high-throughput analyses, and MINT itemizes the number of experiments showing the interaction as well as the evidence for interaction (*e.g.* direct association in a purification, co-localization). Several tools are available to generate and visualize protein interaction networks, of which Cytoscape and Pajek are most commonly used. Cytoscape (25) has the advantage of allowing one to combine information from different large databases such as protein-protein, protein-DNA, and genetic interactions, whereas the Pajek platform can handle large networks with $>10^5$ nodes.

A major challenge in analyzing the medium- to large-scale data sets is in the distinction of true protein interactions from the experimental noise present in the raw data obtained by either large-scale screening method. This necessitates robust methods for the calculation of confidence scores for protein-protein interaction (*i.e.* measures for the likelihood that the observed interaction is true), in particular in the case of qualitative data obtained from yeast two-hybrid and APMS data. Several different confidence scores have been proposed for the systematic evaluation of yeast two-hybrid data sets. For example, Giot *et al.* (8) calculated a confidence score for each measurement using a generalized linear model, which was fit against a training set of known interacting and non-interacting proteins, to determine the cutoff for significant interactions. Braun *et al.* (26) addressed the ambiguity in large-scale protein interaction screens by proposing a dual approach combining a statistical evaluation with follow-up experimentation, which they termed an interaction tool kit. Starting with yeast two-hybrid data, which were evaluated using a logistic regression model, all predicted protein interactions were retested using multiple follow-up assays (26). After each protein interaction was appraised using the tool kit, a probability was generated, indicating if the observed interaction is true (26).

In the case of high-throughput yeast APMS data sets, several computational approaches for calculating a confidence score based on the binary information (*i.e.* a protein is either observed or not observed in the sample) were proposed. For instance, Gavin *et al.* (1) purified hundreds of yeast complexes and defined a "socio-affinity" score based on the frequency of occurrence of all pairs of proteins studied. This score not only takes

the bait-prey relationship into consideration but also regards indirect prey-prey relationships (*i.e.* if two proteins are present in the same purification when a different protein is used as bait) (1). It was suggested that the protein pairs having a high socio-affinity score are more likely to be in direct contact, which could be supported by three-dimensional structures or the data from the yeast two-hybrid system (1). Krogan *et al.* (2) applied a different approach using the Bayesian network to define confidence scores for protein interactions based solely on the bait-prey information. Although the different large-scale data sets are of high quality, some discrepancies exist between individual measurements for the same protein pairs. Therefore, Collins *et al.* (3) proposed an alternative score that incorporates multiple data sets into a single reliable collection of interactions using a novel purification enrichment score. This score takes both positive and negative interactions into consideration, thereby minimizing the effect of false-positive or false-negative information (3). In addition, other methods using a likelihood-based approach or direct graph theory were also applied for the evaluation of qualitative protein-protein interaction data (27). Although each of these scores provides an important source of information for the analysis of protein-protein interactions and the organization of proteins into complexes, no quantitative information is incorporated. As a result, there are emerging efforts to apply quantitative proteomics approaches to the assembly of protein interaction networks.

Quantitative Proteomics and Interaction Networks

The abundance of a peptide can be estimated from the spectra it generates in a mass spectrometer: A highly abundant protein will give rise to mass spectrometric spectra with higher peaks, and in the case of strategies like MudPIT, more spectra will be observed (28). Therefore, this information can be employed to deduce the relative abundance of a protein (28). Initial quantitative strategies were based upon a simultaneous quantification of peptides from two different samples in a single mass spectrometric run (reviewed in Ref. 29). To be able to trace back the origin of the peptides, one of the protein mixtures is stably labeled with a different isotope, which can be distinguished by the mass spectrometer due to its mass difference (reviewed in Ref. 29). One approach to achieve stable isotope labeling is to supply the growth media of the samples to be compared with amino acids composed of different isotopes. For example, in the SILAC (stable isotope labeling by amino acids in cell culture) approach, one of the samples is grown with [^{13}C]arginine (where all six carbon atoms are heavy isotopes); therefore, its peptides can be distinguished from peptides derived from regular [^{12}C]arginine-containing proteins (30). Alternatively, the two samples can also be differently labeled using chemical approaches (reviewed in Ref. 29). Thereby, molecules composed of light *versus* heavy isotopes are covalently incorporated into peptides using reactive groups targeting specific amino acids (*e.g.* thiol-specific reactive groups labeling cysteines or amine-specific reactive groups labeling lysines) (31, 32). In either case, the two differently labeled populations might be mixed before performing the affinity purification, which minimizes experimental variations, or after the affinity purifi-

cation, which avoids cross-binding especially of transient interactions during the purification process.

Label-free quantitative approaches are becoming increasingly popular because they circumvent the time- and cost-demanding process of isotope labeling (reviewed in Ref. 29). Because it was shown that samples from independent MS runs can be employed for the relative quantification of peptides in complex protein mixtures (33), this strategy allows the simultaneous analysis of any number of protein samples. In the case of isotope labeling, this can be achieved only when each sample is run against an internal standard (34). Label-free quantitative APMS experiments require high reproducibility of both the purification and the protein detection steps. To identify and account for such possible variations, Wepf *et al.* (35) suggested the addition of two different isotope-labeled peptides (called SH-quant) before the purification and the MS/MS analysis steps, respectively, which can be used as references during the peptide quantification. Label-free approaches furthermore depend on accurate statistical and bioinformatics tools for a proper comparison as well as on reliable quantitative measures extracted from the spectra. Current scores proposed either evaluate the shape of the peaks (area under the curve or summed intensities) or sum up the total number of unique spectra ("spectral counting") for each peptide (reviewed in Ref. 29).

Recent interaction network analyses have implemented the quantitative information obtained from spectral counting to assign probabilities to individual interactions. In two examples, a general Bayesian approach was applied to determine the posterior probability of an interaction using the spectral count. For instance, Sardiu *et al.* (16, 36) generated a posterior probability between a prey and a bait based on the frequency of occurrence of a protein in a particular bait, whereas Breitkreutz *et al.* (14) introduced the SAINT algorithm, which uses a mixture modeling approach with Bayesian statistical inference to estimate the likelihood of a true interaction. In another case, Sowa *et al.* (18) devised a new metric (D score) that incorporates the uniqueness, the abundance of a prey in a bait, and the reproducibility of each interaction to assign a quantitative score to each prey in every bait. These values are then normalized to a determined threshold value (below which 95% of the data fall), producing a D^N score for each prey in a bait (18). All three methods were successfully applied to construct quantitative protein interaction networks in human (16, 18, 19) and yeast (14, 36). Because the availability of quantitative protein interaction data sets lags behind its binary counterparts, a combination strategy between confidence scores for binary protein-protein interactions and the quantitative scores should be explored to refine the current protein interaction networks.

The drawback of binary protein interaction networks is that every pairwise protein interaction is treated equally (*i.e.* there is either a link or no link). However, depending on the structure and the amino acid composition of the domains responsible for the interaction of two proteins and due to the influence of other players, the strength of interaction between proteins varies. To overcome this limitation, quantitative protein interaction networks can be generated. For quantitative protein interaction networks, each link between two nodes is weighted and corre-

sponds to the quantitative score derived from quantitative APMS experiments (Fig. 1*F*, *left*). Examples for quantitative protein interaction networks in the literature either directly use the quantitative measurements (35) or employ the three different quantitative scores introduced above: that of Sardiu *et al.* (16), the SAINT algorithm (14), and D^N scores (18).

The drawback of using quantitative APMS experiments for the deduction of quantitative protein interaction networks is that the experimental results measured correspond to how two proteins associate in the context of a cell but do not necessary reflect the direct interaction between two proteins. Normally, multiple proteins are co-purified in a single APMS experiment; therefore, the co-purified proteins might and often do influence the interaction of a bait-prey pair. For example, when using baits that are members of stable protein complexes, all purified baits belonging to the same complex will give rise to very similar quantitative values (36). One possibility to overcome this limitation is to perturb the system studied before conducting the quantitative APMS experiments (Fig. 1*F*, *right*). By introducing genetic deletions to different members of the yeast Rpd3 complex, it could be shown that the normally very stable complex dissociates into smaller subcomplexes (36). As expected, these deletions also affected the probabilities of the bait-prey protein pairs derived from these experiments, which therefore are more likely to reflect the direct relationship between bait and prey proteins (36).

Dynamics of Protein Interaction Networks

Biological systems are dynamic; however, the protein interaction networks derived from different large-scale screens are static snapshots of a large population of yeast or human cells. To fully understand dynamic processes such as transcriptional activation or repression, differential protein expression, the cell cycle, or signal transduction, it will be essential to study the effects and consequences of these general processes on local interaction networks of the respective proteins and protein complexes involved. Few studies have been conducted measuring the dynamic changes of protein interaction in the context of some of these processes. Both the human TATA-binding protein complex (37) and the yeast 26 S proteasome complex (21) have been analyzed at different phases of the cell cycle. In addition, von Kriegsheim *et al.* (38) have characterized the changing interacting partners of rat ERK upon its stimulation by epidermal growth factor and nerve growth factor. The analysis of the dynamics of protein interaction networks is expected to increase in the future as protein interaction network analysis tools and technologies become more widely available.

Conclusions

A major hope of biomedical research is to generate models explaining the clinical phenotypes observed in diseases and to develop specific cures. For these efforts, protein interaction networks are widely considered crucial contributors in two different ways: (*a*) by contributing to the characterization and understanding of biological processes and their aberrant function in diseases and (*b*) by providing a framework for the design of specific drugs. Decades of focused experimentation have revealed the significance of local protein interaction networks

in biological processes. Striking examples of such local protein interaction networks are protein complexes, where the combination of multiple proteins into macromolecular complexes leads to functional units that are capable of executing functions that would not be possible with the separate proteins. However, for the majority of protein complexes and protein interactions identified as a result of the large-scale screens conducted, our knowledge is incomplete with respect to their relevance and functional consequences.

Therefore, biomedical research is progressively shifting toward large-scale screens measuring the variables of biological systems (*e.g.* expression of mRNA, protein, and metabolites), their interconnection (*e.g.* protein-DNA and protein-protein interaction), and their functional implications (*e.g.* lethality or genetic interaction screens). To reach the full potential of these screens, the different data sets have to be integrated into large networks modeling single biological processes or whole biological systems. However, these efforts are still in the early stages, and advances in both methodology and bioinformatics analysis are necessary to ever achieve this final goal. In the case of yeast, for which most large-scale data sets are available, the promise of incorporating protein interaction networks into the big picture of biological systems is emerging. For example, it has been observed that proteins that are neighbors in global protein interaction networks are similarly expressed, displaying either related dynamic changes or static expression levels (39). It was furthermore shown that essential proteins have significantly more links in protein networks and are more likely to be hubs (40). Consistent with this observation is that protein pairs that genetically interact (positively or negatively) are more likely to also physically interact (17).

The potential of protein interaction networks in aiding the design of specific drugs was greatly affirmed by a series of studies showing that proteins targeted by the same Food and Drug Administration-approved drugs are more likely to physically interact (41) and that disease genes tend to be hubs in protein interaction networks, exhibiting an elevated number of links (42– 44), therefore making them primary targets for treatments of the respective diseases. Hence, several focused screens or bioinformatics analyses were conducted aiming to generate disease protein networks, *e.g.* for Alzheimer disease (45), Huntington disease (46), Purkinje cell degeneration (47), and breast cancer (48). However, it remains to be seen if drugs can be efficiently designed based upon knowledge derived from protein interaction networks or networks of whole biological systems.

REFERENCES

1. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) *Nature* **440,** 631– 636
2. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrín-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson,

N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006) *Nature* **440,** 637–643

3. Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., Weissman, J. S., and Krogan, N. J. (2007) *Mol. Cell. Proteomics* **6,** 439–450

4. Fields, S., and Song, O. (1989) *Nature* **340,** 245–246

5. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) *Nature* **403,** 623–627

6. Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J. F., Dricot, A., Vazquez, A., Murray, R. R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A. S., Motyl, A., Hudson, M. E., Park, J., Xin, X., Cusick, M. E., Moore, T., Boone, C., Snyder, M., Roth, F. P., Barabási, A. L., Tavernier, J., Hill, D. E., and Vidal, M. (2008) *Science* **322,** 104–110

7. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., and Vidal, M. (2004) *Science* **303,** 540–543

8. Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C. A., Finley, R. L., Jr., White, K. P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R. A., McKenna, M. P., Chant, J., and Rothberg, J. M. (2003) *Science* **302,** 1727–1736

9. Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., and Vidal, M. (2005) *Nature* **437,** 1173–1178

10. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005) *Cell* **122,** 957–968

11. Lievens, S., Lemmens, I., and Tavernier, J. (2009) *Trends Biochem. Sci.* **34,** 579–588

12. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999) *Nat. Biotechnol.* **17,** 1030–1032

13. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) *Nature* **425,** 737–741

14. Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L., Neduva, V., Larsen, B., Lin, Z. Y., Breitkreutz, B. J., Stark, C., Liu, G., Ahn, J., Dewar-Darch, D., Reguly, T., Tang, X., Almeida, R., Qin, Z. S., Pawson, T., Gingras, A. C., Nesvizhskii, A. I., and Tyers, M. (2010) *Science* **328,** 1043–1046

15. Chang, I. F. (2006) *Proteomics* **6,** 6158–6166

16. Sardiu, M. E., Cai, Y., Jin, J., Swanson, S. K., Conaway, R. C., Conaway, J. W., Florens, L., and Washburn, M. P. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 1454–1459

17. Hutchins, J. R., Toyoda, Y., Hegemann, B., Poser, I., Hériché, J. K., Sykora, M. M., Augsburg, M., Hudecz, O., Buschhorn, B. A., Bulkescher, J., Con-

rad, C., Comartin, D., Schleiffer, A., Sarov, M., Pozniakovsky, A., Slabicki, M. M., Schloissnig, S., Steinmacher, I., Leuschner, M., Ssykor, A., Lawo, S., Pelletier, L., Stark, H., Nasmyth, K., Ellenberg, J., Durbin, R., Buchholz, F., Mechtler, K., Hyman, A. A., and Peters, J. M. (2010) *Science* **328,** 593–599

18. Sowa, M. E., Bennett, E. J., Gygi, S. P., and Harper, J. W. (2009) *Cell* **138,** 389–403

19. Behrends, C., Sowa, M. E., Gygi, S. P., and Harper, J. W. (2010) *Nature* **466,** 68–76

20. Malovannaya, A., Li, Y., Bulynko, Y., Jung, S. Y., Wang, Y., Lanz, R. B., O'Malley, B. W., and Qin, J. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 2431–2436

21. Kaake, R. M., Milenković, T., Przulj, N., Kaiser, P., and Huang, L. (2010) *J. Proteome Res.* **9,** 2016–2029

22. Guerrero, C., Milenkovic, T., Przulj, N., Kaiser, P., and Huang, L. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 13333–13338

23. Guerrero, C., Tagwerker, C., Kaiser, P., and Huang, L. (2006) *Mol. Cell. Proteomics* **5,** 366–378

24. Khanin, R., and Wit, E. (2006) *J. Comput. Biol.* **13,** 810–818

25. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) *Genome Res.* **13,** 2498–2504

26. Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J. M., Murray, R. R., Roncari, L., de Smet, A. S., Venkatesan, K., Rual, J. F., Vandenhaute, J., Cusick, M. E., Pawson, T., Hill, D. E., Tavernier, J., Wrana, J. L., Roth, F. P., and Vidal, M. (2009) *Nat. Methods* **6,** 91–97

27. Zhang, B., Park, B. H., Karpinets, T., and Samatova, N. F. (2008) *Bioinformatics* **24,** 979–986

28. Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., and Washburn, M. P. (2006) *J. Proteome Res.* **5,** 2339–2347

29. Kline, K. G., Finney, G. L., and Wu, C. C. (2009) *Brief. Funct. Genomic Proteomic* **8,** 114–125

30. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) *Mol. Cell. Proteomics* **1,** 376–386

31. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) *Nat. Biotechnol.* **17,** 994–999

32. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) *Mol. Cell. Proteomics* **3,** 1154–1169

33. Paoletti, A. C., Parmely, T. J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R. C., Conaway, J. W., Florens, L., and Washburn, M. P. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 18928–18933

34. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100,** 6940–6945

35. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R., and Gstaiger, M. (2009) *Nat. Methods* **6,** 203–205

36. Sardiu, M. E., Gilmore, J. M., Carrozza, M. J., Li, B., Workman, J. L., Florens, L., and Washburn, M. P. (2009) *PLoS One* **4,** e7310

37. Pijnappel, W. P., Kolkman, A., Baltissen, M. P., Heck, A. J., and Timmers, H. M. (2009) *Proteome Sci.* **7,** 46

38. von Kriegsheim, A., Baiocchi, D., Birtwistle, M., Sumpton, D., Bienvenut, W., Morrice, N., Yamada, K., Lamond, A., Kalna, G., Orton, R., Gilbert, D., and Kolch, W. (2009) *Nat. Cell Biol.* **11,** 1458–1464

39. Komurov, K., and White, M. (2007) *Mol. Syst. Biol.* **3,** 110

40. Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., and Gerstein, M. (2004) *Trends Genet.* **20,** 227–231

41. Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabási, A. L., and Vidal, M. (2007) *Nat. Biotechnol.* **25,** 1119–1126

42. Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabási, A. L. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 8685–8690

43. Jonsson, P. F., and Bates, P. A. (2006) *Bioinformatics* **22,** 2291–2297

44. Xu, J., and Li, Y. (2006) *Bioinformatics* **22,** 2800–2805

45. Chen, J. Y., Shen, C., and Sivachenko, A. Y. (2006) *Pac. Symp. Biocomput.* **11,** 367–378

46. Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., Droege, A., Lindenberg, K. S., Knoblich, M., Haenig, C., Herbst, M., Suopanki, J., Scherzinger, E., Abraham, C., Bauer, B., Hasenbank, R., Fritzsche, A., Ludewig, A. H., Buessow, K., Coleman, S. H., Gutekunst, C. A.,

Landwehrmeyer, B. G., Lehrach, H., and Wanker, E. E. (2004) *Mol. Cell* **15,** 853– 865

47. Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabó, G., Rual, J. F., Fisk, C. J., Li, N., Smolyar, A., Hill, D. E., Barabási, A. L., Vidal, M., and Zoghbi, H. Y. (2006) *Cell* **125,** 801– 814

48. Pujana, M. A., Han, J. D., Starita, L. M., Stevens, K. N., Tewari, M., Ahn, J. S., Rennert, G., Moreno, V., Kirchhoff, T., Gold, B., Assmann, V., Elshamy, W. M., Rual, J. F., Levine, D., Rozek, L. S., Gelman, R. S., Gunsalus, K. C., Greenberg, R. A., Sobhian, B., Bertin, N., Venkatesan, K., Ayivi-Guedehoussou, N., Solé, X., Hernández, P., Lázaro, C., Nathanson, K. L., Weber, B. L., Cusick, M. E., Hill, D. E., Offit, K., Livingston, D. M., Gruber, S. B., Parvin, J. D., and Vidal, M. (2007) *Nat. Genet.* **39,** 1338 –1349