Journal of Structural Biology 179 (2012) 320–326

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Novel insights through the integration of structural and functional genomics data with protein networks

Declan Clarke [a], Nitin Bhardwaj [b], Mark B. Gerstein [b,c,d,*]

[a] Department of Chemistry, Yale University, New Haven, CT 06520, USA
[b] Program in Computational Biology and Bioinformatics, Yale University, Bass 426, 266 Whitney Avenue, New Haven, CT 06520, USA
[c] Department of Molecular Biophysics and Biochemistry, Yale University, Bass 432, 266 Whitney Avenue, New Haven, CT 06520, USA
[d] Department of Computer Science, Yale University, 51 Prospect Street, New Haven, CT 06511, USA

ARTICLE INFO

Article history:
Available online 11 February 2012

Keywords:
Protein structures
Protein networks
Hubs
Network prediction and construction
Protein sequence
Evolutionary rate

ABSTRACT

In recent years, major advances in genomics, proteomics, macromolecular structure determination, and the computational resources capable of processing and disseminating the large volumes of data generated by each have played major roles in advancing a more systems-oriented appreciation of biological organization. One product of systems biology has been the delineation of graph models for describing genome-wide protein–protein interaction networks. The network organization and topology which emerges in such models may be used to address fundamental questions in an array of cellular processes, as well as biological features intrinsic to the constituent proteins (or "nodes") themselves. However, graph models alone constitute an abstraction which neglects the underlying biological and physical reality that the network's nodes and edges are highly heterogeneous entities. Here, we explore some of the advantages of introducing a protein structural dimension to such models, as the marriage of conventional network representations with macromolecular structural data helps to place static node and edge constructs in a biologically more meaningful context. We emphasize that 3D protein structures constitute a valuable conceptual and predictive framework by discussing examples of the insights provided, such as enabling *in silico* predictions of protein–protein interactions, providing rational and compelling classification schemes for network elements, as well as revealing interesting intrinsic differences between distinct node types, such as disorder and evolutionary features, which may then be rationalized in light of their respective functions within networks.

1. Background

Biological networks are conventionally represented as maps of nodes and edges, wherein nodes represent biological entities (such as a gene, protein, or miRNA), and edges represent interactions between these entities (such as regulation or physical association). In the case of protein–protein interactions, non-directed edges typically denote physical interactions between pairs of proteins. While many proteins engage in only a small number of interactions, others are highly connected, in that they directly interact with many other proteins. The distinction between nodes with many connections and those with few provides a simple basis for the classification of nodes into two types: hubs and non-hubs, with hubs associating with many binding partners, and non-hubs with few.

In recent years, networks have been studied in greater detail. Early investigations have suggested that many networks (such as the World Wide Web) obey a power law distribution, wherein a node's probability of being associated with k other nodes is proportional to $k^{-\gamma}$ (Barabási and Albert, 1999), resulting in many nodes having low connectivity, and very few nodes connected to many others. Such networks are described as scale-free, as the connectivity distribution does not scale with the total number of nodes (Albert, 2005; Barabási and Albert, 1999). It was later suggested that protein–protein interaction networks obey such power-law distributions (Han et al., 2005; Tanaka et al., 2005; Khanin and Wit, 2006). However, though popular, it must be acknowledged that there is no universal consensus with respect to this conclusion (Khanin and Wit, 2006). Lima-Mendez and van Helden challenge the idea by pointing to statistical and sampling limitations in approaches which have been used to support this concept, and even highlight how this idea has sometimes been supported by flawed practices in plotting degree distribution data (Lima-Mendez and van Helden, 2009).

Nevertheless, the functional roles of network architectures which do exhibit power law distributions may be to confer such networks with a measure of robustness, in that it would render

* Corresponding author at: Department of Computer Science, Yale University, 51 Prospect Street, New Haven, CT 06511, USA.
E-mail address: pi@gersteinlab.org (M.B. Gerstein).

1047-8477/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2012.02.001

their overall integrity more resistant to the random removal of nodes (Barabási and Albert, 1999). Highly-connected nodes and essential genes (i.e., those genes for which knockout results in cell death) have been shown to be correlated (Albert et al., 2000; Jeong et al., 2001; Hahn and Kern, 2005; Maslov and Sneppen, 2002). As an aside, though this observation alone is perhaps not very surprising, it is this very agreement with expectation which suggests that the networks studied may be sufficiently comprehensive to recapitulate intuitive biological phenomena.

Over the past decade, investigators have gained much from newly available data at varying levels of biological organization, including gene expression profiles in different conditions, whole-genome sequences, as well as protein–protein interaction and structural data. Increasingly, this enables the integration of varying forms of information into biological networks, which can help to elucidate known as well as uncover novel features. Our purpose here is to survey some of the ways in which this integration has provided novel insights. Such forms of data include expression correlation, sequence information, and especially the 3D coordinates of the macromolecules and complexes themselves, as provided by X-ray crystallography and NMR. Of course, the structures of the interfaces through which proteins associate constitute essential information about interactions (Keskin et al., 2008a,b). Thus, in particular, we highlight the value of combining 3D structural information with protein interaction data, as evident by providing investigators with powerful means of making predictions about protein–protein interactions, informing the party/date controversy, as well as uncovering properties which are specific to various network node types.

2. Employing structure in the prediction and analysis of network interactions

The importance of how investigators construct high-confidence networks in the first place cannot be overstated, as any subsequent analyses are, of course, contingent on the completeness and accuracy of the map constructed. Many experimental procedures have been devised to detect interactions (Williamson and Sutcliffe, 2010), including yeast two-hybrid assays (Ito et al., 2001) and mass spectrometry coupled with tandem affinity purification (Gavin et al., 2002). Though proven to be of immense value, empirical methods alone can suffer from large numbers of false positive interactions (Deane et al., 2002).

Just as protein structure prediction has received much attention and enthusiasm in the past, an exciting and flourishing discipline in protein interaction network prediction has more recently begun to emerge. The efforts aimed at predicting interactions (and, by extension, entire networks) have taken on a multitude of forms. Given the increasing availability of known monomeric and complex crystal structures, structure-based strategies offer a promising avenue for investigating macromolecular associations, and structural data is increasingly recruited as one of the richest sources of information in the endeavor to taking *in silico* approaches to predicting and understanding the features of nodes and their respective edges.

Toward inferring protein–protein interactions, support vector machines (SVMs) are often adopted, as they have been shown to be extremely powerful in the predicted classification of objects (Boser et al., 1992; Noble, 2006; Ban et al., 2010). For an introduction to SVMs, see the primer by Noble, (2006).

Hue et al. describe an application of SVMs to the protein interaction problem, whereby two known domain structures may be classified as interacting or non-interacting, with the SVM having been trained on pairs of proteins for which interactions are known to exist or known not to exist (Hue et al., 2010). Indeed, this approach is more suitable, in terms of processing power, than predictions based strictly on protein interface docking (Smith and Sternberg, 2002; Inbar et al., 2005), which is far more difficult to apply to the large number of candidates and orientations possible (Grünberg et al., 2006). Machine learning-based approaches may not only provide a means of predicting interactions, but the optimization of SVM-based techniques themselves may shed light on the relative contributions of those physical variables which are most important and conducive to interactions.

The use of SVMs in studying networks has been extended to integral membrane proteins. Miller et al. employ SVMs to assign confidence values to experimentally derived interactions in yeast (Miller et al., 2005), and remark that these results are also noteworthy in that the interactions themselves, of course, better enable biologists to identify the specific functions of membrane proteins, which has been more difficult to accomplish than it is for soluble proteins.

Though the 3D structures of complexes and their constituent proteins continue to be solved at an increasing pace, and despite the more promising prediction capability of using structural data to predict interactions, structural determination has lagged far behind the growing repertoire of genomic sequence data with respect to both volume and rate of production. In addition, the precise 3D characterization of protein interaction pairs is often difficult to obtain experimentally, often because such interactions are transient and unstable to standard experimental procedures. Thus, given the large number of interactions which remain to be structurally defined, increasing attention has been devoted to homology-based approaches for understanding the structural features of those interactions for which X-ray crystallographic data is not yet available (Kiel et al., 2008; Lance et al., 2010; Aloy and Russell, 2002a). Resources based on structural alignment promise to offer even more value as high-throughput structural determination adds to the repertoire of data upon which many of these alignment methods rely (Šali, 1998; Marti-Renom et al., 2000).

Though the strategies by which this may be accomplished vary (Aloy and Russell, 2002a; Lu et al., 2003), many share a similar principle in their general application: generate predicted interactions between a given pair of proteins, domains, or interfaces on the basis of their respective homology to a complex for which 3D experimental data is already available, and then score the fit of the target with the template complex (Aloy and Russell, 2002a). This approach is becoming increasingly promising, given the improved processing resources, more efficient dissemination of sequence data, and the growing pool of structural data available for protein complexes (Aloy and Russell, 2004).

Using structurally defined templates of interacting interfaces, Ogmen et al. have applied structural alignment in order to predict interactions on a target set of proteins, and applied their structural similarity algorithm to build PRISM, a set of publically available analytical tools for prediction, as well as corresponding datasets of putative interactions built from their algorithm's implementation (Ogmen et al., 2005; Aytuna, 2004). Using scores calculated from both structural and evolutionary similarity to the template, an overall confidence score may be assigned to each candidate interaction. In addition, the user may supply their own set of target proteins to exhaustively search for potential interactions.

The confidence of a predicted interaction, as well as the accuracy of the modeled complex structure, are contingent on many factors, including the resolution of the known 3D template structure(s), the degree of sequence homology between the target and the template pairs, the degree of disorder in the interfaces themselves, and the force field used in refining the final modeled complex (Kiel et al., 2008). In addition, the rotamer library (Mendes et al., 1999) used to properly configure the orientations of amino acids belonging to the targets is important in generating a reasonable model of the interaction (Kiel et al., 2008).

Though promising as a means of probing networks at the level of structure, homology-based interaction modeling is not without limitations. A major challenge will be achieving the accuracy demanded by the specificity of many interactions. That is, even if the general features of a protein's architecture can be hammered down, it may yet prove difficult to achieve interface complementarily for the many interactions for which even slight changes in interface geometry and chemistry result in ablated interactions. It is also difficult to estimate the thermodynamic properties of the resulting modeled structure (such as affinity), as such properties are contingent on the specific amino acids and their respective orientations within the interface (Kiel et al., 2008), as well as more global physical properties of the constituent proteins. In addition, homologous pairs of proteins may interact in different ways (Aloy and Russell, 2002b; Aloy et al., 2003; Kim et al., 2006b), and homology in the protein fold itself may not always reliably be used to predict interactions (Aloy et al., 2003).

Along these same lines, although homology-based approaches have achieved some success, it is far more difficult to model the actual contributions of single amino acids to binding energies. Protein interfacial hotspots are amino acids or clusters of amino acids which contribute disproportionately large values to interaction energies, and are thus critical for binding and specificity (Bogan and Thorn, 1998). Although determining the contributions of single amino acids to interaction binding affinities may be performed experimentally (often by measuring the effects on binding through single residue mutagenesis of individual residues to alanine), computational approaches to this problem may provide the same information with significantly less time and effort. This is facilitated by the fact that hotspots are generally characterized by unique sequential and biophysical properties. Solvent accessibility and inter-residue potentials have been used to predict hot spots with considerable reliability (Tuncbag et al., 2009a). Of course, the successful prediction of hot spots not only holds relevance for protein interaction networks, but also for drug development (Bogan and Thorn, 1998), wherein the design of therapeutic agents may be motivated and guided by knowledge of the residues most responsible for interactions. Targeting such residues with high-affinity binding compounds provides a direct way to interfere with those protein interactions which constitute pathways that are most responsible for the progression of disease states.

3. The party/date hub dilemma: A case study of insights through structure

3.1. Background

Within the hub category, it may be possible to further classify and analyze nodes on the basis of features evident from gene expression data (Han et al., 2004; Luscombe et al., 2004), subcellular localization (Han et al., 2004), or protein structural characteristics (Kim et al., 2006a, 2008). On the basis of expression correlation among the binding partners of hubs, Han et al. introduced the notion that two fundamentally different hub types exist (Han et al., 2004). Those hubs for which higher (lower) expression correlation values exist between the hub's binding partners were categorized as "party" ("date") hubs. Another line of evidence in support of this distinction was the differing effects on network architecture upon the deletion of party or date hubs, as well as the greater localization entropy observed for date relative to party binding partners (Han et al., 2004).

The concept of a biological module (a set or subsystem of closely interacting proteins which, together, function as a unit to carry out a specific biological role; Hartwell et al., 1999) is central to the different functionalities ascribed to these hub types; it is argued that party hubs function within modules by interacting with several partners, which are present simultaneously, in order to carry out their biological roles, whereas date hubs function by interconnecting modules (Han et al., 2004), as the higher expression correlation values and lower localization entropy for party hub partners result if each partner, along with the hub, must be present simultaneously and in the same place to function as a module, whereas date hubs would interact with modules which operate differently in time and space. They note that such a model would also explain the differing network integrity effects upon deleting party and date hubs: the removal of party hubs would not have as much of an effect as the removal of core date hubs.

Batada et al. published work (Batada et al., 2006) questioning the conclusions reached in the analyses described above. Their expanded network was much more tolerant to hub deletion, and the network integrity (as measured by the largest subnetwork remaining upon hub deletion) was maintained for both party and date hub deletion. Batada et al. also question the bimodality of the expression patterns used to define party and date hubs. In addition, Batada et al. reason that date hubs may have partners of higher localization entropy as a consequence of the fact that nodes classified as date hubs generally have more binding partners. Finally, Batada et al. find no significant differences in evolutionary rates between party and date hubs.

3.2. The party/date debate in light of protein structure

Here, we discuss the party/date hub debate simply as an example of how the consideration of distinct physical interfaces can shed light on an existing problem in network biology. Kim et al. introduced the structural interaction network (SIN), in which edges are structurally annotated on the basis of sequence homology to structurally defined complexes, and subsequently employed three-dimensional structural exclusion to define distinct interfaces on each protein (Kim et al., 2006a; Kim et al., 2008). The interactions which constitute the SIN are considered mutually exclusive if they involve a common interface of a particular protein, and are otherwise classified as compatible. Hubs (i.e., those nodes with at least five interaction partners) were classified as *singlish-* or *multi-interface hubs* (those with less than or at least three distinct interfaces, respectively). Statistically significant disparities were observed between these hub types in terms of essentiality, co-expression, and evolutionary rate.

Multi-interface proteins, which are considered to be capable of simultaneous interactions, are marked by higher expression correlation with their binding partners, and the characterization of singlish and multi-interface hubs helps to explain the party/date hub model, with multi-interface hubs being more similar to party hubs, and singlish–interface hubs more like date hubs (Kim et al., 2006a). Indeed, singlish–interface hubs, with only one or two distinct interfaces, would be unable to interact with a large number of partners simultaneously, and it would thus be intuitively reasonable if the proteins which bind to singlish–interface hubs are expressed under different cellular contexts, growth stages, or subcellular localizations–that is, the proteins to which they bind would have higher entropy values for these properties (Kim et al., 2006a). Many of the multi-interface hubs likely evolved their interaction interfaces in order to meet the need to simultaneously interact with many partners, and so it is likely that such partners are expressed and localized in relative unison (Kim et al., 2006a). Fig. 1 provides a schematic of this idea, in which different colors denote different biological modules (as adapted from Fig. 1a of Han et al., 2004), and the central white node represents a singlish interface hub capable of mutually exclusive interactions with these modules. Kim et al. point out that the expression correlation for singlish-interface hubs with their binding partners is 0.17, whereas for multi-interface hubs, this value was 0.25,

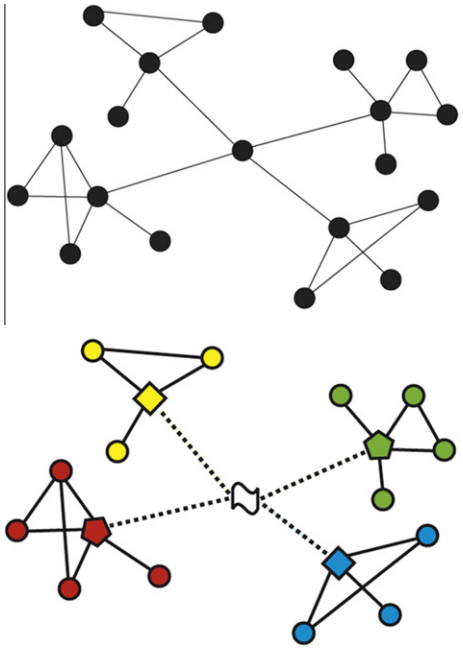

Fig.1. Correspondence between singlish-interface (multi-interface) and date (party) hubs. Simplified schematics of a network are shown. *Top*: a simple node-and-edge network, as derived from binary protein–protein interaction data alone. *Bottom*: The network which results when structural data on interfaces and interactions is applied to the simple node-and-edge network above. In this rendering, n-sided polygons represent nodes with n distinct interfaces, and circles are generic representations of nodes. Different colors represent distinct biological modules. The central white node is a singlish interface hub capable of interacting with multiple different modules. Solid lines denote simultaneously possible interactions, and dashed lines represent mutually exclusive interactions. Multi-interface hubs make up the cores of each of the four modules shown. Thus, this figure represents a correspondence between singlish- or multi-interface hubs and date or party hubs, respectively.

further supporting this relationship between the number of interfaces and party/date hub class.

Though the debate over the party/date distinction has not been entirely resolved (Agarwal et al., 2010; Kahali et al., 2009; Cukuroglu et al., 2010; Vallabhajosyula et al., 2009; Patil et al., 2010), the consideration of physical interfaces adds greater richness and dimension to the story, and lends some support to the significance of this distinction.

3.3. Disorder

Many proteins contain long sequences which fail to adopt well-defined structures (Uversky, 2002), and these elements are frequently described as intrinsically disordered (Dunker et al., 2002; Linding et al., 2003; Iakoucheva et al., 2004; Radivojac et al., 2007). It is now appreciated that intrinsically disordered elements play essential roles in basic protein functionality, such as interaction with other cellular species (Dyson and Wright, 2005). Phosphorylation sites, which are frequently the gate points in signaling pathways, tend to be surrounded by disordered elements (Iakoucheva et al., 2004). It is natural to ask what roles disorder might play in the context of networks. Do hub proteins exhibit different levels of disorder than those with low degree? Do multi- and singlish-interface hubs exhibit differing levels of disorder, and if so, how might such differences be rationalized? For those categories of network nodes which exhibit a greater degree of disorder, what roles might disorder play for such node types?

Naturally, these and related questions have been difficult to address by purely structure-based experimental means; by their very nature, disordered elements do not lend themselves to crystallization, for instance. As a result, many approaches aimed at identifying disordered regions rely on sequence data instead (Romero et al., 1997a; Obradovic et al., 2003).

In addition to disparities in the sequence complexity of ordered and disordered segments (Romero et al., 2001), there are differences between some of the physical properties of their amino acids (Dunker et al., 2001, 1998; Romero et al., 1997a,b, 1998). Disordered sequences tend to be enriched in charged or polar residues (Romero et al., 2001; Vucetic et al., 2003; Dunker et al., 2005). This is intuitively reasonable if disordered sequences explore large fields of conformational space in solution, and more ordered elements are buried by a tightly packed protein environment. These and other differences enable one to use SVMs for predicting disorder (Hue et al., 2010; Noble, 2006).

With respect to protein interaction networks, the potential functions of intrinsically disordered elements are discussed and reviewed by Dunker et al. (Dunker et al., 2005). Here, the authors describe several potential roles of disordered elements. One may be the ability of disordered sequences to confer high-degree nodes with a greater degree of binding promiscuity. Disordered sequences may constitute highly flexible links between more structurally defined domains, thereby conferring such linked domains with a much greater degree of mobility relative to one another, which would enable binding to more geometrically diverse species.

Using proteome-wide data available from *Homo sapiens* (Rual et al., 2005; Stelzl et al., 2005), *Saccharomyces cerevisiae* (Uetz et al., 2000; Ito et al., 2001), *Drosophila melanogaster* (Giot et al., 2003), and *Caenorhabditis elegans* (Li et al., 2004), Haynes et al. examined the disorder in hubs relative to minimally connected nodes (Haynes et al., 2006). Note that, in this study, hubs were much more conservatively defined as those proteins which interact with at least ten partners, and non-hubs were defined as those with only one interaction. Using several sets of criteria for describing disorder, hubs were found to exhibit a much higher degree of intrinsic disorder than did minimally connected nodes. The authors also examined the potential relationship between disorder and protein biological function using gene ontology categories in yeast, and observed consistent disparities for different categories.

Building on their previous work in structural network analysis, Kim et al. analyzed the relative disorder of different network node types in greater detail (Kim et al., 2008). The sequence-based prediction of disordered elements was performed by applying DISOPRED (Ward et al., 2004) to nearly 7000 ORFs in yeast. Consistent with the findings described above (Haynes et al., 2006), hub proteins were found to be more disordered than the proteome average. One interesting finding was the observation that only singlish hubs were found to exhibit this disparity; multi-interface hubs did not differ significantly from the remainder of the proteome. The authors explain the greater degree of disorder observed for singlish hubs by pointing out that these nodes are more likely to bind to one another, as seen, for instance, in cell signaling cascades, which generally tend to constitute more disordered elements (Iakoucheva et al., 2002). In addition, the set of binding interfaces for both hub types were found to be well ordered. The authors report that the a typical singlish hub's interface binds to a greater number of domains than does a multi-interface hub, and they explain this phenomenon by observing that the binding partners of singlish hubs are disordered relative to the remainder of the proteome, and this disorder is partially responsible for enabling binding to a more physically diverse set of proteins.

4. Evolutionary rates in the context of structural networks

Data obtained through genome-wide analyses employing next-generation sequencing technology (Metzker, 2010) may be used to learn about the evolutionary patterns that emerge in networks, which grow primarily through a combination of gene duplications, thereby adding nodes to the network, and nonsynonymous single nucleotide polymorphisms, thereby potentially rewiring the existing network by changing the specificity of protein interfaces (Berg et al., 2004). By combining duplications with mutations responsible for changing interactions, networks which reconstitute biological attributes may be generated *in silico* (Berg et al., 2004; Wagner, 2003). This ability to recapitulate global statistical features characteristic of biological networks (such as connectivity distributions) lends support to the notion that the evolution of network topology is largely driven by duplications and interface mutations, and that sequence data may tell much of the story not only of the evolution of the nodes themselves, but also of network organization (Shou et al., 2011; Xia et al., 2009; Fraser et al., 2002; Bloom and Adami, 2003; Lemos et al., 2005; Wagner, 2001; Yu et al., 2004).

Evolutionary rates may be determined by examining the ratio of nonsynonymous to synonymous base pair substitutions (dN/dS). The structural interaction network built by Kim et al. has shown a strong relationship between dN/dS ratios and the number of interfaces (Kim et al., 2006a). Singlish-interface hubs are generally faster-evolving than multi-interface hubs, which evolve slower than the remainder of the proteome; more specifically, this relationship is a consequence of the close correspondence between dN/dS ratios and the proportion of a hub's surface area that is involved in interactions with other proteins (Kim et al., 2006a, 2008). This relationship between evolutionary rate and interaction surface area had also been reported previously (Fraser and Hirsh, 2004). Importantly, this trend puts the evolutionary dynamics of network nodes in a more direct, structurally meaningful context: the hub interfaces themselves (for both singlish and multi-interface hubs) are slow-evolving, with the greater number of interfaces in multi-interface hubs likely contributing to their overall lower evolutionary rates (Kim et al., 2006a).

The findings described above are in agreement with the observation that singlish-interface hubs also tend to be less ordered, as higher degrees of disorder may enable faster evolutionary changes (Brown et al., 2002). The relationship between disorder and evolutionary rates was noted by (Kim et al., 2008), and this phenomenon is perhaps to be expected, largely because of the tolerance that disordered and loosely-packed elements would have to amino acid changes.

In related work employing similar approaches, proteins within the network periphery have been found to be under positive evolutionary selection; in contrast, proteins more central to networks are more evolutionarily constrained (Kim et al., 2007). The structural explanation for this relationship is much the same as that discussed above: sites within interfaces tend to be under negative evolutionary selection, and given that protein interfaces have been shown to be generally more conserved than other surface residues (Teichmann, 2002), these results are to be expected. Indeed, interfaces are, in some respects, similar to protein cores in that stringent geometric and biophysical constraints are usually imposed by the need to closely interact with complementary sets of amino acids. Future efforts may further be directed toward a more careful examination of the disparity between the evolutionary rates of interface and core residues, and more specifically, how these disparities vary depending on hub and interaction types.

5. Conclusion and future directions

The networks constructed to represent biological activity in yeast, humans, and other organisms are far from complete. The information contained therein provides only a fraction of the true interactomes of these organisms, though it has been difficult to pin down the degree to which network representations capture the full repertoire of associations. In terms of both graph network renderings and 3D structural definitions of interactions, there is much which remains to be added. The quality of these and future analyses are, as discussed, contingent on how thoroughly the modeled networks reflect the full complement of interactions present in living cells. However, as noted, the trends and findings surveyed here lend support to the integrity of current network models.

It is likely that, in building network representations, much will be gained through a combination of the experimental and *in silico* approaches. Some have already begun to move in this direction. For instance, homology-based approaches have been combined with empirical methods, such as high-resolution microscopy and tandem affinity purification, to study multi-subunit complexes, including the exosome and RNA polymerase II (Aloy et al., 2004). In addition, degrees of redundancy in the interactions reported by different approaches may be used to assign confidence measures to macromolecular associations (Kiel et al., 2008).

The insights gained through the recently added structural dimension have taught us a great deal, and further work in the area of structural networks will pave the way for the addition of yet greater dimension to network organization and behavior. Tuncbag et al. have analyzed networks with an eye toward time (Tuncbag et al., 2009b). Here, the authors point out the need to model networks by referring to the mutually exclusive nature with which multiple binding partners interact with a hub having only a finite number of interfaces. They underscore that static representations of networks can be made into more biologically realistic constructs (specifically, as processes) with the introduction of time. Given that this analysis is so dependent on defining the structures of constituent nodes, follow-up work of a similar nature would benefit greatly from an expanded structure network.

A more comprehensive structural definition of networks also paves the way for yet another layer of information: that of molecular motions. Bhardwaj et al. have taken advantage of solved structures of alternative conformations of nodes in the construction of a type of dynamic structure interaction network, termed DynaSIN (Bhardwaj et al., 2011). Here, network topology was combined with structural information in order to elucidate the potential relationships between protein structural modularity and node type. It was found that, in general, there is a positive relationships between the number of interfaces and potential degrees of conformational change, as measured by RMSD.

New forms of data are becoming increasingly recognized for their value in gaining novel insights into basic network biology, and the volume of this data (especially genomic sequence data) continues to grow rapidly. Next-generation sequencing technology greatly facilitates the measurement of gene expression across the entire genome in a variety of conditions, and an understanding of co-regulated genes better enables the investigator to infer interactions. Though it fails to keep pace with sequence information, available structural data in the PDB (Berman et al., 2000) is also growing at a considerable pace. In addition, *in silico* approaches have become better at inferring physical interactions. Over the course of the next several years, the careful integration of these and other forms of data should provide investigators with networks that are morebiologically meaningful, in both time and space.

Acknowledgments

D.C. acknowledges the support of the NIH Predoctoral Program in Biophysics (T32 GM008283-24). N.B. and M.B.G. acknowledge support from the NIH.

References

Agarwal, S., Deane, C.M., Porter, M.A., Jones, N.S., 2010. Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. PLoS Comput. Biol. 6 (6), 000817.
Albert, R., Jeong, H., Barabási, A.L., 2000. Error and attack tolerance of complex networks. Nature 406, 378–382.
Albert, R., 2005. Scale-free networks in cell biology. J. Cell Sci. 118, 4947–4957.
Aloy, P., Russell, R.B., 2002a. Interrogating protein interaction networks through structural biology. Proc. Natl. Acad. Sci. USA 99, 5896–5901.
Aloy, P., Russell, R.B., 2002b. The third dimension for protein interactions and complexes. Trends Biochem. Sci. 27, 633–638.
Aloy, P., Ceulemans, H., Stark, A., Russell, R.B., 2003. The relationship between sequence and interaction divergence in proteins. J. Mol. Biol. 332, 989–998.
Aloy, P., Russell, R.B., 2004. Ten thousand interactions for the molecular biologist. Nature Biotechnol. 22, 1317–1321.
Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A.C., Bork, P., Superti-Furga, G., Serrano, L., et al., 2004. Structure-based assembly of protein complexes in yeast. Science 303, 2026–2029.
Aytuna,A.S., 2004. A high performance algorithm for automated prediction of protein–protein interactions. Master thesis, Graduate School of Engineering, Koc University, Istanbul,Turkey.
Ban, H.J., Heo, J.Y., Oh, K.S., Park, K.J., 2010. Identification of type 2 diabetes-associated combination of SNPs using Support Vector Machine. BMC Genet. 11, 26.
Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.
Batada, N.N., Hurst, L.D., Tyers, M., 2006. Evolutionary and physiological importance of hub proteins. PLoS Comput. Biol. 2 (7), e88.
Berg, J., Lässig, M., Wagner, A., 2004. Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. BMC Evol. Biol. 4, 51.
Berman, H.M., Westbrook, J., Feng, Z., et al., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.
Bhardwaj, N., Abyzov, A., Clarke, D., Shou, C., Gerstein, M.B., 2011. Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions. Protein Sci. 20, 1745–1754.
Bloom, J.D., Adami, C., 2003. Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein–protein interactions data sets. BMC Evol. Biol. 3, 21.
Bogan, A.A., Thorn, K.S., 1998. Anatomy of hot spots in protein interfaces. J. Mol. Biol. 280, 1–9.
Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Haussler, D. (Ed.), 5th Annual ACM Workshop on COLT. ACM Press, Pittsburgh, PA, pp. 144–152.
Brown, C.J., Takayama, S., Campen, A.M., Vise, P., Marshall, T.W., Oldfield, C.J., Williams, C.J., Dunker, A.K., 2002. Evolutionary rate heterogeneity in proteins with long disordered regions. J. Mol. Evol. 55, 104–110.
Cukuroglu, E., Ozkirimli, E., Keskin, O., 2010. Structural properties of hub proteins.
Deane, C.M., Salwinski, L., Xenarios, I., Eisenberg, D., 2002. Two methods for assessment of the reliability of high throughput observations. Mol. Cell. Proteomics 1, 349–356.
Dunker, A.K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., Villafranca, J.E., 1998. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac. Symp. Biocomput., 473–484.
Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., et al., 2001. Intrinsically disordered protein. J. Mol. Graph Model 19, 26–59.
Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., Obradovic, Z., 2002. Intrinsic disorder and protein function. Biochemistry 41, 6573–6582.
Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., Uversky Vladimir, N., 2005. Flexible nets: The roles of intrinsic disorder in protein interaction networks. FEBS J. 272, 5129–5148.
Dyson, H.J., Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell. Biol. 6, 197–208.
Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., Feldman, M.W., 2002. Evolutionary rate in the protein interaction network. Science 296, 750–752.
Fraser, H.B., Hirsh, A.E., 2004. Evolutionary rate depends on number of protein–protein interactions independently of gene expression level. BMC Evol. Biol. 4, 13.
Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G., 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147.
Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., et al., 2003. A protein interaction map of Drosophila melanogaster. Science 302, 1727–1736.
Grünberg, R., Nilges, M., Leckner, J., 2006. Flexibility and conformational entropy in protein-protein binding. Structure 14, 683–693.
Hahn, M.W., Kern, A.D., 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol. Biol. Evol. 22, 803–806.
Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., et al., 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430, 88–93.
Han, J.D., Dupuy, D., Bertin, N., Cusick, M.E., Vidal, M., 2005. Effect of sampling on topology predictions of protein-protein interaction networks. Nat. Biotechnol. 23, 839–844.
Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W., 1999. From molecular to modular cell biology. Nature 402, C47–C52.
Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., Iakoucheva, L.M., 2006. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput. Biol. 2, e100.
Hue, M., Riffle, M., Vert, J.P., Noble, W.S., 2010. Large-scale prediction of protein-protein interactions from structures. BMC Bioinformatics 11, 114.
Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., Dunker, A.K., 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. J. Mol. Biol. 323, 573–584.
Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z., Dunker, A.K., 2004. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32, 1037–1049.
Inbar, Y., Benyamini, H., Nussinov, R., Wolfson, H.J., 2005. Prediction of multimolecular assemblies by multiple docking. J. Mol. Biol. 349, 435–447.
Ito, T. et al., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98, 4569–4574.
Jeong, H., Mason, S.P., Barabási, A.L., Oltvai, Z.N., 2001. Lethality and centrality in protein networks. Nature 411, 41–420.
Kahali, B., Ahmad, S., Ghosh, T.C., 2009. Exploring the evolutionary rate differences of party hub and date hub proteins in *Saccharomyces cerevisiae* protein–protein interaction network. Gene 429, 18–22.
Khanin, R., Wit, J., 2006. How scale-free are biological networks. J. Comput. Biol. 13, 810–818.
Keskin, O., Gursoy, A., Ma, B., Nussinov, R., 2008a. Principles of protein–protein interactions: what are the preferred ways for proteins to interact? Chem. Rev. 108, 1225–1244.
Keskin, O., Tuncbag, N., Gursoy, A., 2008b. Characterization and prediction of protein interfaces to infer protein–protein interaction networks. Curr. Pharm. Biotechnol. 9, 67–76.
Kiel, P., Beltrao, L., Serrano, 2008. Analyzing protein interaction networks using structural information. Annu. Rev. Biochem. 77, 415–441.
Kim, P.M., Lu, L.J., Xia, Y., Gerstein, M.B., 2006a. Relating three-dimensional structures to protein networks provides evolutionary insights. Science 314 (5807), 1938–1941.
Kim, W.K., Henschel, A., Winter, C., Schroeder, M., 2006b. The many faces of protein–protein interactions: a compendium of interface geometry. PLoS Comput. Biol. 2, 1151–1164.
Kim, P.M., Korbel, J.O., Gerstein, M.B., 2007. Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. Proc. Natl. Acad. Sci. USA 104, 20274–20279.
Kim, P.M., Sboner, A., Xia, Y., Gerstein, M., 2008. The role of disorder in interaction networks: a structural analysis. Mol. Syst. Biol. 4, 179.
Lance, B.K., Deane, C.M., Wood, G.R., 2010. Exploring the potential of template-based modelling. Bioinformatics 26, 1849–1856.
Lemos, B., Bettencourt, B.R., Meiklejohn, C.D., Hartl, D.L., 2005. Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein–protein interactions. Mol. Biol. Evol. 22, 1345–1354.
Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., et al., 2004. A map of the interactome network of the metazoan C. elegans. Science 303, 540–543.
Lima-Mendez, G., van Helden, J., 2009. The powerful law of the power law and other myths in network biology. Mol. Biosyst. 5 (12), 1482–1493.
Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., 2003. Protein disorder prediction: implications for structural proteomics. Structure 11, 1453–1459.
Lu, L., Arakaki, A.K., Lu, H., Skolnick, J., 2003. Multimeric threading-based prediction of protein–protein interactions on a genomic scale: application to the *Saccharomyces cerevisiae* proteome. Genome Res. 13, 1146–1154.
Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., Gerstein, M., 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312.
Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Šali, A., 2000. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291–325.
Maslov, S., Sneppen, K., 2002. Specificity and stability in topology of protein networks. Science 296, 910–913.
Mendes, J., Baptista, A.M., Carrondo, M.A., Soares, C.M., 1999. Improved modeling of side-chains in proteins with rotamer-based methods: a flexible rotamer model. Proteins 37, 530–543.
Metzker, M.L., 2010. Sequencing technologies - the next generation. Nat. Rev. Genet., 11.

Miller, J.P., Lo, R.S., Ben-Hur, A., Desmarais, C., Stagljar, I., Noble, W.S., Fields, S., 2005. Large-scale identification of yeast integral membrane protein interactions. Proc. Natl. Acad. Sci. USA 102, 12123–12128.
Noble, W.S., 2006. What is a support vector machine? Nat. Biotechnol. 24, 1565–1567.
Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K., 2003. Predicting intrinsic disorder from amino acid sequence. Proteins 53 (Suppl. 6), 566–572.
Ogmen, U., Keskin, O., Aytuna, A.S., Nussinov, R., Gursoy, A., 2005. PRISM: protein interactions by structural matching. Nucleic Acids Res. 33, W331–W336.
Patil, A., Kinoshita, K., Nakamura, H., 2010. Hub promiscuity in protein-protein interaction networks. Int. J. Mol. Sci. 11 (4).
Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N., Dunker, A.K., 2007. Intrinsic disorder and functional proteomics. Biophys. J. 92, 1439–1456.
Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J.E., Dunker, A.K., 1997a. Identifying disordered regions in proteins from amino acid sequence. Proc. Int. Conf. Neural Networks 1, 90–95.
Romero, P., Obradovic, Z., Dunker, A.K., 1997b. Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inform. Ser. Workshop Genome Inform. 8, 110–124.
Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E., Garner, E., Guilliot, S., Dunker, A.K., 1998. Thousands of proteins likely to have long disordered regions. Pac. Symp. Biocomput., 437–448.
Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., Dunker, A.K., 2001. Sequence complexity of disordered protein. Proteins 42, 38–48.
Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., et al., 2005. Towards a proteome-scale map of the human protein–protein interaction network. Nature 437, 1173–1178.
Šali, A., 1998. 100,000 protein structures for the biologist. Nat. Struct. Biol. 5, 1029–1032.
Shou, C., Bhardwaj, N., Lam, H.Y., Yan, K.K., Kim, P.M., Snyder, M., Gerstein, M.B., 2011. Measuring the evolutionary rewiring of biological networks. PLoS Comput. Biol. 7, e1001050.
Smith, G.R., Sternberg, M.J., 2002. Prediction of protein–protein interactions by docking methods. Curr. Opin. Struct. Biol. 12, 28–35.
Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., et al., 2005. A human protein–protein interaction network: a resource for annotating the proteome. Cell 122, 957–968.
Tanaka, R., Yi, T.M., Doyle, J., 2005. Some protein interaction data do not exhibit power law statistics. FEBS Lett. 579, 5140–5144.
Teichmann, S.A., 2002. J. Mol. Biol. 324, 399–407.
Tuncbag, N., Gursoy, A., Keskin, O., 2009a. Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics 25, 1513–1520.
Tuncbag, N., Kar, G., Gursoy, A., Keskin, O., Nussinov, R., 2009b. Towards inferring time dimensionality in protein-protein interaction networks by integrating structures: the p53 example. Mol. Biosyst. 5, 1770–1778.
Uversky, V.N., 2002. What does it mean to be natively unfolded? Eur. J. Biochem. 269, 2–12.
Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., et al., 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403, 623–627.
Vallabhajosyula, R.R., Chakravarti, D., Lutfeali, S., Ray, A., Raval, A., 2009. Identifying hubs in protein interaction networks. PLoS ONE 4, e5344.
Vucetic, S., Brown, C.J., Dunker, A.K., Obradovic, Z., 2003. Flavors of protein disorder. Proteins 52, 573–584.
Wagner, A., 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Mol. Biol. Evol. 18, 1283–1292.
Wagner, A., 2003. How the global structure of protein interaction networks evolves. Proc. Biol. Sci. 270, 457–466.
Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., Jones, D.T., 2004. The DISOPRED server for the prediction of protein disorder. Bioinformatics 20, 2138–2139.
Williamson, M.P., Sutcliffe, M.J., 2010. Protein-protein interactions. Biochem. Soc. Trans. 38 (4), 875–878.
Xia, Y., Franzosa, E.A., Gerstein, M.B., 2009. Integrated assessment of genomic correlates of protein evolutionary rate. PLoS Comput. Biol. 5 (6), e1000413.
Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., Gerstein, M., 2004. Genomic analysis of essentiality within protein networks. Trends Genet. 20, 227–231.