# Topological properties of protein interaction networks from a structural perspective

**Attila Gursoy\*, Ozlem Keskin\* and Ruth Nussinov†‡[1]**

\*Center for Computational Biology and Bioinformatics and College of Engineering, Koc University, Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey, †Basic Research Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI-Frederick, Frederick, MD 21702, U.S.A., and ‡Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

## Abstract

Protein–protein interactions are usually shown as interaction networks (graphs), where the proteins are represented as nodes and the connections between the interacting proteins are shown as edges. The graph abstraction of protein interactions is crucial for understanding the global behaviour of the network. In this mini review, we summarize basic graph topological properties, such as node degree and betweenness, and their relation to essentiality and modularity of protein interactions. The classification of hub proteins into date and party hubs with distinct properties has significant implications for relating topological properties to the behaviour of the network. We emphasize that the integration of protein interface structure into interaction graph models provides a better explanation of hub proteins, and strengthens the relationship between the role of the hubs in the cell and their topological properties.

## Introduction

Cellular functions are co-ordinated activities of many proteins and biological molecules such as DNA, RNA and ligands that interact with each other. A system of interacting elements can be abstracted with a mathematical structure called a graph. A basic graph is a collection of nodes (vertices) and edges. In most studies of biological networks, system elements such as proteins and genes are represented by graph nodes and the physical interactions between elements are represented by edges. Edges can be directed, identifying the source and target nodes in the interaction. Directed edges can be useful for interactions such as phosphorylation, where the source node is the protein to phosphorylate, the edge is the phosphorylation process, and the target is the same protein in the phosphorylated state. Here, we will review topological and structural properties of large interaction networks, particularly the role of hub proteins (highly connected proteins) and the importance of structural information (protein interfaces) in the characterization of hubs.

Protein interaction networks are commonly represented by undirected graphs (graphs whose edges have no direction); hence, we will stay with undirected graph representations. Graphs can be characterized by several properties. The number of edges incident on a node, called the degree, is one of the basic properties. A molecule interacting with many other distinct molecules would have a high node degree; for example, ATP interacts with many distinct proteins. A simple path is a sequence of distinct, connected nodes in a graph. In signal transduction networks, a path could represent a sequence of interactions from the sensing of a signal to its target; for example, a molecule binding to a receptor eventually leading to a transcription factor binding to DNA. The shortest path between two nodes is the path with minimum length (usually length is the number of edges). The average path length (also known as characteristic path length) of a graph is computed by averaging over all shortest paths between all pairs of nodes. It is an important network statistic and represents *closeness* and consequently how quickly information can be transferred in a network. Most real world networks have unexpectedly short characteristic path lengths, as popularized with the six degrees of freedom play. This property is called *small world* property and is studied in detail in [1,2]. Figure 1(B) displays an example of a yeast protein–protein interaction network, and a small subgraph illustrating a hub protein (node A in Figure 1A).

During the past decade, with the increasing availability of data on real world networks, numerous studies have focused on large-scale topological network properties. The influential work by Barabasi and Albert [3] initiated the interest in topological properties of biological networks. In their work, the World Wide Web was mapped to a graph, where web pages were the nodes and the links between the pages were the edges. The number of links of a node was observed to follow a power law distribution, that is, the probability of a node having degree $k$ is proportional to $k^{-\gamma}$, and the distribution is independent of the number of nodes; hence these networks are called scale free. Scale-free networks have many nodes with small degrees and allow nodes with high degrees (hubs) with decreasing probability. Many (but not all, see [4–6]) biological networks exhibit scale-free behaviour. An extensive review of scale-free networks in biology is given in [7]. Scale-free networks are resistant to random node failures. Even if a significant number of randomly selected

**Figure 1 | Protein interaction networks**
(**A**) A simple graph model of a protein interaction network. Node A is a hub, node B is a non-hub (a high-betweenness node). (**B**) A sample protein interaction network derived from yeast protein interactions.
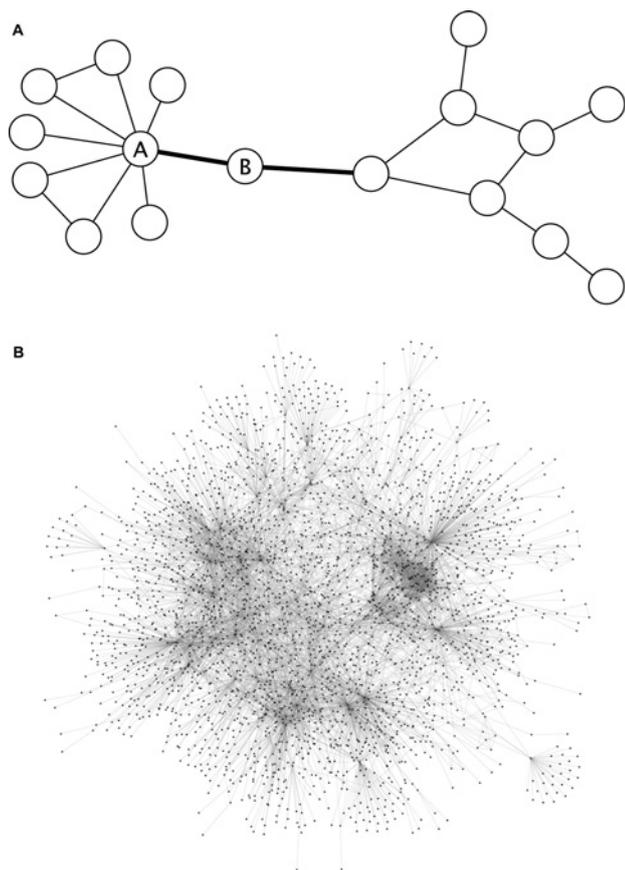


nodes are removed from the network, the remaining nodes will still be connected [8]. Random removal will take out many small-degree nodes, there are many of them; however, a few remaining highly connected hubs will still preserve the connectivity of the network. On the other hand, scale-free networks are vulnerable to targeted attacks to the hubs. Jeong et al. [9] conducted a series of protein–protein interaction network simulations to study the impact of failures and attacks on protein networks. The network contained 1870 protein nodes and 2240 physical interactions gathered from yeast. The degree distribution of the network was consistent with power law, characterizing it as scale free. Large-scale phenotypic (gene knockout) experiments, have indicated that yeast can tolerate many single gene deletions. In accord with the experiments, simulation of random node deletion did not cause topological changes; on the other hand, removal of highly connected nodes increased the network diameter (defined as characteristic path length). Highly connected nodes causing network failures and *essential* genes (i.e. genes whose knockout leads to unviable cells) were observed to be correlated, a property known as *lethality–centrality*. The degree centrality has been used to characterize the node importance in several other studies as well [10–12]. Yu et al. [12]

pointed out that most essential genes in regulatory networks correspond to house-keeping genes whose protein products perform basic functions; hence they must always function. Although degree distribution and the gene essentiality are found to be related, the degree distribution alone may not be sufficient to characterize the network since the same number of connections can be realized in various ways [13]. In fact, some recent works [14–16] report that correlations might be the result of bias in available interaction networks.

*Betweenness* centrality is yet another global property of networks [17]. By specifying how a node influences the communication between two nodes, it is possible to locate important, though not very highly connected network nodes. Figure 1(A) shows an example of high betweenness (node B). The betweenness of a node is defined as the ratio of the number of shortest paths passing through a node to the total number of paths passing through the nodes. Considering the modular and hierarchical architecture of biological networks [18–20], betweenness can be important for finding non-hub important nodes [21,22] or classifying hubs according to their positions in the network.

## Intermodule hubs

Investigations aimed at distinguishing hubs from non-hub proteins increasingly uncover topological properties of the hubs in the protein–protein interaction networks. In many studies, typically, protein interaction networks were viewed as static structures, collapsed in time, due to their complexity and the limited availability of temporal data in biological interactions. In reality, interactions have both space and time dimensions. Integrating microarray mRNA expression profiles, which provide valuable temporal information on protein–protein interactions, into network models allows us to incorporate the time dimension. The motivation behind the integration is based on the assumption that products of co-regulated genes have higher chances of interacting with each other.

This assumption can be used to improve the prediction of cellular pathways. For example, Segal et al. [23] integrated mRNA expression profiles and protein–protein interaction data using a probabilistic graph model, resulting in more meaningful group of genes that might be part of a common pathway.

Time information might explain how hub proteins influence the network structure in space and time as well. Han et al. [19] integrated static high-quality yeast protein–protein interaction data with a compilation of mRNA microarray expression levels to characterize the hub proteins. The interaction data had 1379 proteins with an average degree of 3.6. The degree distribution revealed a power-law relationship exhibiting scale-free network architecture. The temporal characteristic of hub proteins is determined by computing the correlation of their mRNA expressions with the expressions for their partner proteins. Two types of hub proteins were identified. The ones with relatively high correlations between their mRNA levels and their partners,

called party hubs, were assumed to be interacting with their partners concurrently. The hubs with low correlation, on the other hand, were assumed to interact with their partners at different times, and they were called date hubs.

Party hubs are thought to have a local role in the network, with strong connections within functional modules; on the other hand, date hubs interconnect functional modules giving rise to the network topology, and consequently have a global role. For example, the date hub Cmd1 connects modules related to homoeostasis of cations, protein folding and stabilization, budding, and endoplasmic reticulum. On the other hand, the interactions of the party hub Vti1 are all in one module, that of the endoplasmic reticulum [19]. Systematic removal of hubs without distinguishing between party or date hubs was observed to be more harmful to the network connectivity than random protein removal in previous studies. However, Han et al. [19] proposed that the network failure was due to attacks on the date hubs, not on party hubs. When the network is attacked only at party hubs, the characteristic path length did not differ from random failures. On the other hand, the network started to disintegrate and the characteristic path length increased with attacks on date hubs. This is possible, for example, if the network has a modular structure and date hubs interconnect the modules [18]. Although date hubs are those primarily responsible for the increase in network diameter and disintegration, both hub types were observed to have similar essentiality (i.e. removal of a hub is likely to be lethal). A party hub can be indispensable for a critical function, thus scoring high as an essential protein. However, it is argued that date hubs assist in the co-ordination between functional modules and hence perturbation of date hubs might have a different impact on the network than that of party hubs.

## Essential interactions

An alternative explanation to the centrality–lethality rule was proposed by He and Zhang [14]. According to this explanation, some interactions are more essential, and the functional importance of hubs is not due to their high degree, but rather to certain interactions. An essential protein–protein interaction is defined to be an interaction indispensable to the survival of an organism. Most network models treat all interactions as equally important. Differentiating interactions and treating them differently might provide improved interpretation of topological properties. Since experimental determination of essential protein–protein interactions is more difficult than determination of the essential proteins, He and Zhang have used a computational approach to relate essential protein–protein interactions to essential proteins and investigated their implications. Based on the assumption that an essential interaction makes the two proteins (involved in the interaction) essential, they estimated the number of essential interactions as a function of the IBEP (interaction between essential proteins). The yeast protein–protein interaction network manually compiled by the Yeast Comprehensive Genome Database with 4126 proteins and

7356 edges has been used to find IBEPs. An IBEP does not have to be an essential interaction; but the number of essential interactions in a network is assumed to increase in proportion to the number of IBEPs. By repeated random rewiring of the interactions without changing the degree of the nodes, a distribution of the number of IBEPs in the network has been computed. The average calculated number of IBEPs was much lower than the observed IBEPs in the protein–protein interaction network, suggesting that the additional IBEPs in the real network are the outcome of the essential interactions. Analysis of the yeast protein–protein interaction network further suggested that the centrality–lethality rule might be attributed not only to the relative importance of hubs, but also to essential interactions. Hubs tend to have a higher chance of having essential interactions since they have many connections; hence their removal is expected to be more lethal as compared with non-hubs. The significance of essential interactions is supported by the observation that party and date hubs have similar essentiality [19], but their removal will have a different impact on the network connectivity: while removal of party hubs does not have an impact on the global network connectivity, removal of date hubs disintegrates the network. The lethality of party hubs might be due to the loss of some of their essential interactions. The study has suggested that different edges (interactions) in the network have different levels of importance. Treating these interactions accordingly, may lead to more potent drug targets.

## Proteins with HBLC (high betweenness but low connectivity)

Several studies focusing on the betweenness in proteins shed light on a non-hub centric organization (co-ordination of modules by non-hub proteins) of the protein interaction networks and in a way supported the idea that some interactions are more important in the network structure. Valente and Cusick characterized the yeast proteome using a betweenness centrality and modularity score (the ratio of the numbers of intra-module edges/total edges) [24]. The co-ordinated functionality of the yeast interactome is likely to be achieved not only by hub proteins but also by the pairwise interactions of connectors. These connector nodes are high traffic nodes (i.e. they have high betweenness scores) mediating co-ordination across modules. Comparing a protein's betweenness score and its essentiality among similar connectivity proteins did not show a significant correlation. Turning off one connector protein mediating between two distinct modules might not affect the functioning of the individual modules. One may speculate that attacking protein interaction networks (by drugs) on one node might not be enough [25], and knowledge of the connector proteins and the modular topology may provide information regarding which combination of nodes need to be attacked.

In the growth model proposed by Barabasi and Albert [3], a new node forms interactions with other nodes with a probability proportional to the degree of the target node (preferential attachment), leading to a scale-free topology. Preferential

attachment by itself is not based on processes promoting the growth of biological networks, hence fails to model modularity properties observed in such networks. The main biological processes that contribute to the growth (evolutionary dynamics) of biological networks are point mutations and gene duplications. Growth models based on the gene-duplication growth add proteins randomly with their interactions duplicated. To imitate mutations of duplicated genes, some existing links are rewired [26,27]. Valente and Cusick [24] report that the gene-duplication model fails as well to produce a non-hub-centric modular organization of the yeast interactome. On the other hand, introduction of a process called link dynamics that simulates gain or loss of interactions due to mutations in addition to gene duplication [26,27] resulted in the formation of module structures in the generated network.

Proteins with low connectivity but high betweenness may play a key role in the modular structure of the yeast interactome. A study by Joy et al. [21] using a graph theoretical betweenness of the yeast protein interaction data from the recent DIP and MIPS (2005) interaction databases showed that HBLC nodes are abundant in the yeast proteome and suggests a modular structure, with some proteins outside the highly connected modules co-ordinating the information flow between the modules. HBLC nodes are not explained either by the growth model of Barabasi and Albert [3] or by gene-duplication alone [28]. However, the link dynamics model [26] resulted in the formation of HBLC structures in the generated network. Link dynamics corresponds to mutations in protein interfaces which affect certain interactions, which is a biologically more realistic process than random introduction of nodes. Separation of mutation and gene duplication rates are based on the fact that, on the evolutionary scale, mutational events leading to gain or loss of certain protein–protein interactions are faster than gene duplication. To support the significance of HBLC nodes, in a very recent study by Yu et al. [22], non-hub bottleneck nodes (HBLC nodes) are shown to be more essential, and betweenness is found to be a more significant indicator of essentiality than degree.

## Subgraph centrality and essential proteins

Real biological networks also have a high clustering coefficient [29], i.e. the neighbours are likely to be connected. Network motifs found in biological networks are small subgraphs that represent a specific communication module or process [30,31]. Estrada and Rodríguez-Velázquez proposed taking motifs, or subgraphs, into consideration as a measure of centrality [32]. Subgraph centrality characterizes nodes by considering long-range effects due to participation in structurally nearby subgraphs. This is the major difference between a subgraph centrality and other centrality measures: the influence of a node by participation in small subgraphs. A comparison of subgraph centrality with other centrality measures for the yeast protein interaction network is reported in [32]. The ranking of proteins by degree centrality and subgraph cent-

rality revealed that more essential proteins are ranked higher with subgraph centrality. Frequent participation in small structural motifs rather than the number of direct interactions makes a protein indispensable (centrality–lethality relation).

## Structural analysis of hubs and interfaces in protein interaction networks

The hub-centric view of modularity and the party and date hub proteins distinction have been questioned recently by Batada et al. [33]. Using multi-validated HC (high-confidence) yeast interaction data, Batada et al. [33] reported that hub–hub interactions are not suppressed, in contrast with the results of [11,19]. In addition, the evidence for date and party hub distinction cannot be instantiated within the HC dataset. The contradicting results are probably due either to biased data (Y2H data) [34] or, although more accurate, still incomplete (HC) data [35].

Most of the work studying the protein interaction networks did not consider structural information at the atomic resolution level (protein interfaces). Protein interfaces provide valuable insight on how proteins interact with each other [36–40] and can be used to improve the prediction of protein interactions, for example [41–44]. One of the early attempts to combine three-dimensional structure information of proteins with interactions in the context of systems biology includes the work of Aloy and Russell [45–47]. They used knowledge of structures to interpret the interactions identified by two-hybrid methods. Some interactions were found to be indirect (mediated by other proteins) by comparing interacting proteins and their participation in three-dimensional complexes.

By using structural information in the analysis of hub proteins, Kim et al. [48] clarified some of the questions related to the distinction between date and party hubs. Proteins from a (HC) high consensus yeast interaction network are mapped to known structures using Pfam (Protein Families Database of Alignments and Hidden Markov Models; http://www.sanger.ac.uk/Software/Pfam/) domains [49]. Only those interactions that are contained in protein complexes are retained in the final network. The structures of complexes are used to find interface regions on the protein surfaces. An interaction is classified as mutually exclusive if the interface it uses is shared by another interaction; otherwise it is classified as simultaneous. This approach allowed them to classify the hub proteins as: (i) singlish-interface hubs interacting with several other proteins using only one shared binding surface, which corresponds to date hubs; and (ii) multi-interface hubs corresponding to party hubs (interacting with many proteins simultaneously). They report that multi-interface hubs evolve more slowly than the others. The multi-interface hubs are found to be more closely correlated with their partners' expression levels compared with singlish-interface hubs. These results are in agreement with the previous studies discriminating hub proteins [19,50]. Figure 2 illustrates the usage of interfaces to distinguish between date and party hubs.

**Figure 2 | Use of interfaces to indicate simultaneous and unshared interactions**

Circle nodes represent proteins, diamonds indicate their interfaces; S for shared, U for unshared interface. Node A corresponds to a party hub (has many unshared interfaces for simultaneous interactions). Node B is a date hub. The shared interface is used for different interactions at different times.
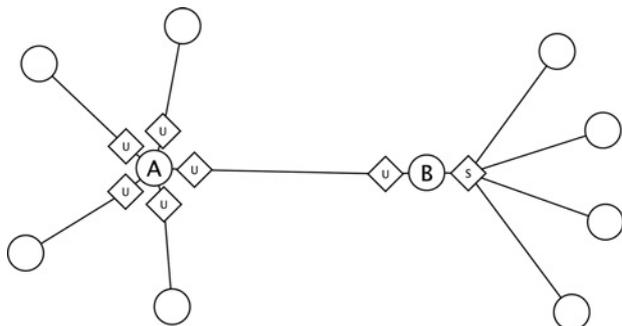


## Conclusions

With the availability of an enormous amount of diverse biological data, including complete genome sequences (e.g. the human genome project) and protein structures (with the Protein Data Bank growing exponentially), computational methods have become increasingly successful in inferring protein–protein interactions. Microarray experiments investigate the expression of thousands of genes. Knowledge of co-regulated genes further allows us to infer interactions. While there are false positives and false negatives within this massive amount of information, these advances, nevertheless, lead to a huge amount of interaction data. Systems biology seeks to organize this information to allow quantitative modelling of the dynamics of complex cellular pathways and facilitate comprehension of how cells are regulated.

Graph theory can be applied successfully to biological networks to relate the structure of the network to biological functions. Protein interactions are commonly represented by undirected graphs. One of the basic properties of nodes, degree, has been extensively used to characterize the proteins, namely hub proteins. Hub proteins have been observed to be related to essential genes. Increasing interest and recent work on network analysis of protein interaction data revealed, however, that degree by itself, although very informative, is not sufficient to relate network properties of hubs to their function. For example, hub proteins have been classified as date and party hubs by using additional data such as microarray expression profiles. Betweenness of nodes in a graph seems to be another very important property. Some important but non-hub nodes (HBLC proteins) were related to proteins mediating co-ordination between modules.

Using only the network of protein interaction data, however, does not support these findings strongly, and one needs to be careful in drawing conclusions relating network structure to biological function. The source of the problem is the incompleteness or bias in protein interaction data coming from experimental and computational predictions. However, incorporation of structural data (protein structure and interfaces) into networks allowed us to re-enforce the previous findings and further provided a better explanation of hub proteins. The analysis of hub proteins by considering their interfaces (single- or multi-interface hubs) together with important interactions (HBLC nodes) can have important implications on modelling protein interaction networks and evaluating drug design targeting protein–protein interactions such multi-target drugs.

## References

1 Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. Nature **393**, 440–442
2 Strogatz, S.H. (2001) Exploring complex networks. Nature **410**, 268–276
3 Barabasi, A.L. and Albert, R. (1999) Emergence of scaling in random networks. Science **286**, 509–512
4 Han, J.D., Dupuy, D., Bertin, N., Cusick, M.E. and Vidal, M. (2005) Effect of sampling on topology predictions of protein–protein interaction networks. Nat. Biotechnol. **23**, 839–844
5 Tanaka, R., Yi, T.M. and Doyle, J. (2005) Some protein interaction data do not exhibit power law statistics. FEBS Lett. **579**, 5140–5144
6 Khanin, R. and Wit, E. (2006) How scale-free are biological networks. J. Comput. Biol. **13**, 810–818
7 Albert, R. (2005) Scale-free networks in cell biology. J. Cell Sci. **118**, 4947–4957
8 Albert, R., Jeong, H. and Barabasi, A.L. (2000) Error and attack tolerance of complex networks. Nature **406**, 378–382
9 Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature **411**, 41–42
10 Hahn, M.W. and Kern, A.D. (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol. Biol. Evol. **22**, 803–806
11 Maslov, S. and Sneppen, K. (2002) Specificity and stability in topology of protein networks. Science **296**, 910–913
12 Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X. and Gerstein, M. (2004) Genomic analysis of essentiality within protein networks. Trends Genet. **20**, 227–231
13 Friedel, C.C. and Zimmer, R. (2007) Influence of degree correlations on network structure and stability in protein–protein interaction networks. BMC Bioinformatics **8**, 297
14 He, X. and Zhang, J. (2006) Why do hubs tend to be essential in protein networks? PLoS Genet. **2**, e88
15 Batada, N.N., Hurst, L.D. and Tyers, M. (2006) Evolutionary and physiological importance of hub proteins. PLoS Comput. Biol. **2**, e88
16 Coulomb, S., Bauer, M., Bernard, D. and Marsolier-Kergoat, M.C. (2005) Gene essentiality and the topology of protein interaction networks. Proc. Biol. Sci. **272**, 1721–1725
17 Goh, K.I., Oh, E., Jeong, H., Kahng, B. and Kim, D. (2002) Classification of scale-free networks. Proc. Natl. Acad. Sci. U.S.A. **99**, 12583–12588
18 Fraser, H.B. (2005) Modularity and evolutionary constraint on proteins. Nat. Genet. **37**, 351–352

19 Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P. et al. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature **430**, 88–93

20 Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. and Friedman, N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat. Genet. **34**, 166–176

21 Joy, M.P., Brock, A., Ingber, D.E. and Huang, S. (2005) High-betweenness proteins in the yeast protein interaction network. J. Biomed. Biotechnol. **2**, 96–103

22 Yu, H., Kim, P.M., Sprecher, E., Trifonov, V. and Gerstein, M. (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput. Biol. **3**, e59

23 Segal, E., Wang, H. and Koller, D. (2003) Discovering molecular pathways from protein interaction and gene expression data. Bioinformatics **19**, i264–i271

24 Valente, A.X. and Cusick, M.E. (2006) Yeast protein interactome topology provides framework for coordinated functionality. Nucleic Acids Res. **34**, 2812–2819

25 Csermely, P., Agoston, V. and Pongor, S. (2005) The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol Sci. **26**, 178–182

26 Berg, J., Lässig, M. and Wagner, A. (2004) Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. BMC Evol. Biol. **4**, 51

27 Wagner, A. (2003) How the global structure of protein interaction networks evolves. Proc. Biol. Sci. **270**, 457–466

28 Pastor-Satorras, R., Smith, E. and Sole, R.V. (2003) Evolving protein interaction networks through gene duplication. J. Theor. Biol. **222**, 199–210

29 Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. Nat. Rev. Genet. **5**, 101–113

30 Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002) Network motifs: simple building blocks of complex networks. Science **298**, 824–827

31 Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat. Genet. **31**, 64–68

32 Estrada, E. and Rodríguez-Velázquez, J.A. (2005) Subgraph centrality in complex networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **71**, 056103

33 Batada, N.N., Reguly, T., Breitkreutz, A. and Boucher, L. (2006) Stratus not altocumulus: a new view of the yeast protein interaction network. PLoS Biol. **4**, e317

34 Aloy, P. and Russell, R.B. (2002) Potential artefacts in protein-interaction networks. FEBS Lett. **530**, 253–254

35 Hakes, L., Pinney, J.W., Robertson, D.L. and Lovell, S.C. (2008) Protein–protein interaction networks and biology–what's the connection? Nat. Biotechnol. **26**, 69–72

36 Jones, S. and Thornton, J.M. (1996) Principles of protein–protein interactions. Proc. Natl. Acad. Sci. U.S.A. **93**, 13–20

37 Keskin, O. and Nussinov, R. (2007) Similar binding sites and different partners: implications to shared proteins in cellular pathways. Structure **15**, 341–354

38 Kim, W.K., Henschel, A., Winter, C. and Schroeder, M. (2006) The many faces of protein–protein interactions: a compendium of interface geometry. PLoS Comput. Biol. **2**, e124

39 Keskin, O., Gursoy, A., Ma, B. and Nussinov, R. (2008) Principles of protein–protein interactions: what are the preferred ways for proteins to interact? Chem. Rev. **108**, 1225–1244

40 Keskin, O., Tuncbag, N. and Gursoy, A. (2008) Characterization and prediction of protein interfaces to infer protein–protein interaction networks. Curr. Pharm. Biotechnol. **9**, 67–76

41 Aloy, P. and Russell, R.B. (2003) InterPreTS: protein interaction prediction through tertiary structure. Bioinformatics **19**, 161–162

42 Aytuna, A.S., Gursoy, A. and Keskin, O. (2005) Prediction of protein–protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics **21**, 2850–2855

43 Ogmen, U., Keskin, O., Aytuna, A.S., Nussinov, R. and Gursoy, A. (2005) PRISM: protein interactions by structural matching. Nucleic Acids Res. **33**, W331–W336

44 Beltrao, P., Kiel, C. and Serrano, L. (2007) Structures in systems biology. Curr. Opin. Struct. Biol. **17**, 378–384

45 Aloy, P. and Russell, R.B. (2006) Structural systems biology: modelling protein interactions. Nat. Rev. Mol. Cell Biol. **7**, 188–197

46 Aloy, P. and Russell, R.B. (2002) Interrogating protein interaction networks through structural biology. Proc. Natl. Acad. Sci. U.S.A. **99**, 5896–5901

47 Aloy, P. and Russell, R.B. (2002) The third dimension for protein interactions and complexes. Trends Biochem. Sci. **27**, 633–638

48 Kim, P.M., Lu, L.J., Xia, Y. and Gerstein, M.B. (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. Science **314**, 1938–1941

49 Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) The Pfam protein families database. Nucleic Acids Res. **30**, 276–280

50 Ekman, D., Light, S., Bjorklund, A.K. and Elofsson, A. (2006) What properties characterize the hub proteins of the protein–protein interaction network of *Saccharomyces cerevisiae*? Genome Biol. **7**, R45