We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



20

Protein-Protein Interaction Networks: Structures, Evolution, and Application to Drug Design

Takeshi Hase[1,2] and Yoshihito Niimura[3]
[1]*Department of Bioinformatics, Graduate School of Biomedical Science, Tokyo Medical and Dental University, Tokyo,*
[2]*The Systems Biology Institute, Tokyo,*
[3]*Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo,*
Japan

1. Introduction

Since proteins exert their functions through interaction with other proteins rather than in isolation, networks of protein interactions are inevitable for understanding protein functions, disease mechanisms, and discovering novel targets of therapeutic drugs (Hase et al. 2009, Barabasi et al. 2011, Vidal et al. 2011). With the recent influx of genome-wide data of protein interactions, many researchers have studied on the structures and statistics of protein-protein interaction networks (PINs). To discover novel drug target genes, it is informative to understand topological and statistical characteristics of PINs, and how disease and drug target genes are distributed over the networks. Moreover, because those statistical properties of PINs are the results of long-term evolution, analysis of the PIN architecture from the viewpoint of comparative genomics and molecular evolution is of particular importance.

In this chapter, we will first summarize our current knowledge of the statistical properties of PINs. We then argue on possible evolutionary mechanisms generating those properties and review the studies related to drug discovery and diseases as an application of the analyses of PIN structure. Finally, we briefly discuss the possibilities of medical studies as an integration of network and evolutionary biology.

2. Genome-wide data of protein-protein interactions

Genome-wide protein-protein interaction data have been obtained from several organisms, including *Escherichia coli* (Arifuzzaman et al. 2006), *Saccharomyces cerevisiae* (Uetz et al. 2000, Ito et al. 2001, Guldener et al. 2006, Reguly et al. 2006, Yu et al. 2008), *Plasmodium falciparum* (LaCount et al. 2005), *Arabidopsis thaliana* (Arabidopsis Interactome Mapping Consortium 2011), *Caenorhabditis elegance* (Li et al. 2004, Simonis et al. 2009), *Drosophila melanogaster* (Giot et al. 2003), and *Homo sapiens* (Rual et al. 2005, Stelzl et al. 2005). Table 1 summarizes the PIN

datasets that are currently available. These data were mainly obtained by high-throughput experimental techniques such as yeast two-hybrid (Y2H) screens and tandem affinity purification followed by mass spectrometry (APMS) screens (Deane et al. 2002, Parrish et al. 2006, Lavallee-Adam et al. 2011), as well as extensive literature curation by experts.

Species	Number of proteins	Number of interactions	Data type	References
Mycoplasma pneumoniae	410	1,058	APMS	Kuhner et al. (2009)
MRSA 252	608	13,219	APMS	Cherkasov et al. (2011)
Treponema pallidum	726	3,649	Y2H	Titz et al. (2008)
Mesorhizobium loti	1,804	3,121	Y2H	Shimoda et al. (2008)
Escherichia coli	2,448	8,625	APMS	Arifuzzaman et al. (2006)
Campylobacter jejuni	1,301	11,557	Y2H	Parrish et al. (2007)
Yeast	1,647	2,518	Y2H	Yu et al. (2008)
	3,224	11,291	Literature curated	Reguly et al. (2006)
	3,891	7,270	Manually curated	MIPS
	3,278	4,549	Y2H	Ito et al. (2001)
	1,004	957	Y2H	Uetz et al. (2000)
Malaria parasite	1,267	2,726	Y2H	LaCount et al. (2005)
Arabidopsis thaliana	2,661	5,664	Y2H	Arabidopsis Interactome Mapping Consortium (2011)
Worm	2,898	5,240	Y2H	Li et al. (2004)
	2,528	3,864	Y2H	Simonis et al. (2009)
Fly	4,679	4,780	Y2H	Giot et al. (2003)
	2,477	3,546	Y2H	Pacifico et al. (2006)
Human	2,783	6,007	Y2H, Literature curated	Rual et al. (2005)
	1,613	3,101	Y2H	Stelzl et al. (2005)

Table 1. PIN datasets. Y2H, Yeast two-hybrid screens; APMS, tandem affinity purification followed by mass spectrometry screens. "Manually curated" indicates that interactions obtained from high-throughput screens and literatures are manually integrated by experts.

Y2H screens examine an interaction between two proteins, by expressing these genes in yeast nucleus as fusion proteins (Parrish et al. 2006). One protein is fused to a DNA-binding domain of a transcription factor (*e.g.*, Gal4 and LexA), and the other protein is fused to a transcription

activation domain of the transcription factor. When two proteins interact with each other, DNA-binding domain and activation domain are indirectly connected. The activation domain can then interact with the transcription start site of the reporter genes (*e.g.*, LacZ). From the expression of the reporter gene, the interaction between two proteins can be detected. In APMS screens, affinity purification selectively purifies a protein complex that includes a protein of interest (bait protein) (Lavallee-Adam et al. 2011). Then, from the purified complex, mass spectrometry identifies possible interacting partners of the bait protein.

It has been pointed out that genome-wide PIN data identified by high-throughput experiments contains a large number of false positive interactions (Hakes et al. 2008). Y2H screens may detect possible interactions between two proteins that actually reside in different subcellular localizations (Deane et al. 2002). APMS studies identify many false positive interactions caused by inadequate purification (Lavallee-Adam et al. 2011).

Literature-curated PIN datasets are likely to be more reliable, because interactions included in such datasets were obtained from small-scale experiments. However, those data are derived from hypothesis-driven researches focusing on several proteins that are supposed to be biologically important, and thus the datasets can be highly biased (Arabidopsis Interactome Mapping Consortium 2011). Therefore, to study the global structure of PINs, researchers should use several datasets obtained by various methods.

3. Statistical properties of PINs

In PINs, a protein and a physical interaction between two proteins are represented as a node and a link, respectively. A series of studies have revealed that PINs have several interesting properties from the viewpoint of network architecture.

3.1 Scale-freeness

The number of links for a given node is called a degree. The degree distribution $P(k)$, the fraction of nodes with k degrees in a network, has been used to characterize the global structure of a network.

Erdös and Renyi (1960) investigated a random network with N nodes, in which links are attached between each pair of nodes with a uniform probability p. This network contains approximately $pN(N\text{-}1)/2$ randomly placed links. Erdös and Renyi (1960) showed that, in a random network, the distribution $P(k)$ follows the Poisson distribution (Fig 1A, left). Therefore, most nodes have degrees that are nearly equal to the mean degree $<k>$ among all nodes in the network.

On the other hand, the distribution $P(k)$ of various technological, social, and biological networks including PINs is known to follow the power law, *i.e.*, $P(k) \sim k^{-\gamma}$ (Albert et al. 1999; Fig 1A, right). These networks are highly heterogeneous; they have a large number of low-degree nodes and a small but significant number of high-degree nodes that are called hubs. A network following the power law does not have a typical degree characterizing most nodes in the network (*e.g.*, the mean degree $<k>$ in a random network), and thus it is called a "scale-free" network. It was shown that scale-free networks are very robust against random removal of nodes, although selective removal of hubs drastically changes their structures (Jeong et al. 2001, Han et al. 2004).

3.2 Small-worldness

The cluster coefficient of nodes *i* is defined as $C_i = 2e_i/k_i(k_i\text{-}1)$, where k_i is the degree of node *i*, and e_i is the number of links among k_i neighbors of node *i* (Watts & Strogatz 1998) (see Fig. 1B). In other words, e_i is the number of triangles that pass through node *i*. C_i is equal to one when all neighbors of node *i* fully interact with one another, while C_i is 0 when there are no links among the neighbors of node *i*. The mean of the cluster coefficient among all nodes, <*C*>, reflects the density of triangles ("cliques") within a network.

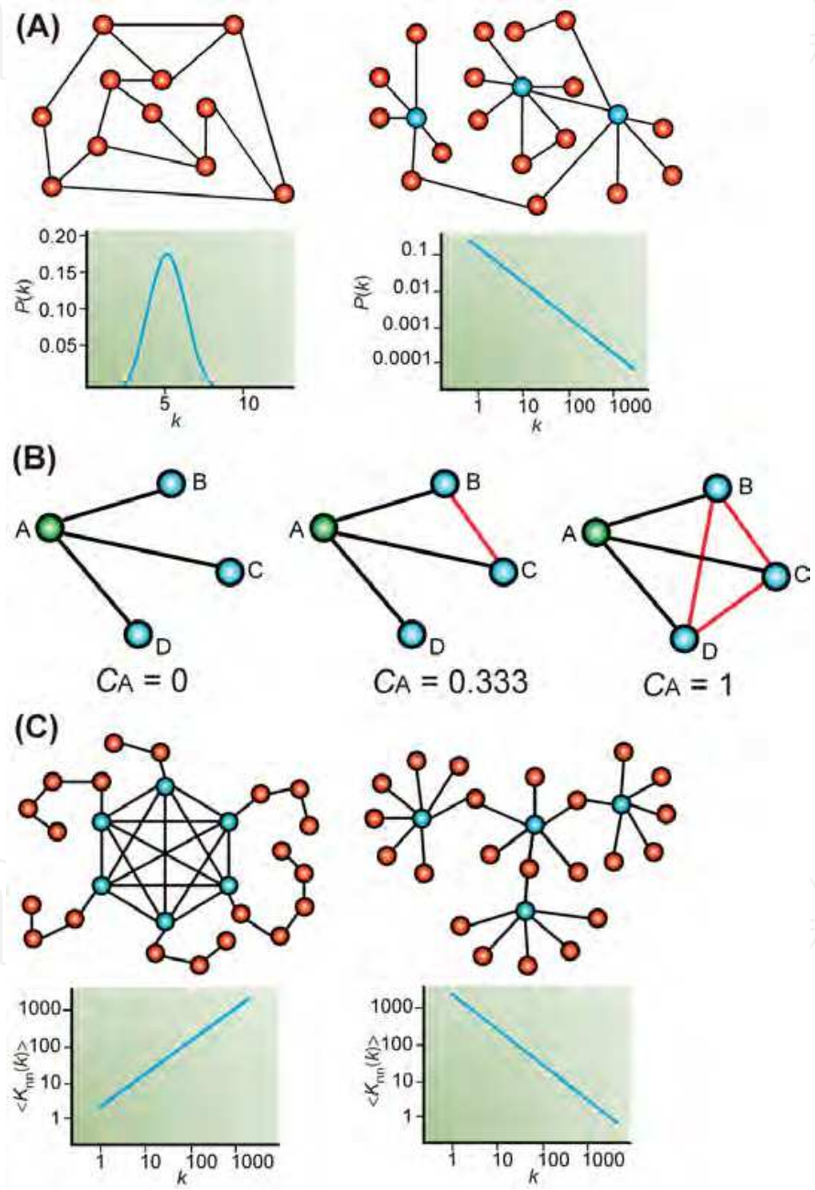


Fig. 1. Measures of a network structure
(A) A random network (left) and a scale-free network (right). The degree distribution $P(k)$ is shown below the networks. **(B)** Cluster coefficient. Red lines represent links among three neighbors of node A. The numbers of links (e_A) among nodes B, C, and D (the neighbors of

node A) in the left, middle, and right networks are 0, 1, and 3, respectively. The cluster coefficient C_A of node A is shown for each network. **(C)** Assortative (left) and disassortative (right) networks. The distribution of $<K_{nn}(k)>$ is shown below the networks. Blue and Red nodes indicates hubs and non-hubs, respectively.

The shortest path length between a pair of nodes is the smallest number of links (distance) that are necessary for travelling from one node to the other (Barabasi & Oltvai 2004). The mean shortest path length among all possible pairs of nodes in a network is denoted by $<L>$. Watts and Strogatz (1998) found that a random network has a much smaller value of $<L>$ compared with a regular lattice. Based on this observation, they defined a "small-world" network as a network that has a value of $<L>$ as small as a random network but is highly clustered like a regular lattice. In a random network, $<L> \sim \log N/\log<k>$, and $<C> = <k>/N$, where N is the number of nodes.

In PINs, the value of $<L>$ is small and the value of $<C>$ is much higher than a random network; therefore, PINs are generally considered to be small-world networks. However, several studies showed that PINs are actually "ultra-small", because $<L>$ is considerably smaller than that in a random network (Chung & Lu 2002, Cohen & Havlin 2003, Hase et al. 2008). In a PIN, proteins are located close to each other, suggesting that perturbations given to a single protein would affect the behaviour of many other proteins and even the entire PIN.

3.3 Assortativity

Another statistic characteristic of a network is the correlation between degrees of nodes that are linked to each other (Callaway et al. 2001, Newman 2002, Costa et al. 2007). Pearson correlation coefficient r of the degrees at both ends of a link is used to evaluate the degree correlation. Networks with $r > 0$ and $r < 0$ are called as assortative and disassortative networks, respectively. In an assortative network, hubs tend to be connected to each other (Fig 1C, left), while in a disassortative network, hubs tend to have links to low-degree nodes (Fig 1C, right).

$<K_{nn}(k)>$, the mean degree among the neighbors of all k-degree nodes ("nn" in $<K_{nn}(k)>$ means "nearest neighbors"), is also used to evaluate the assortativity of a network (Pastor-Satorras et al. 2001, Maslov & Sneppen 2002, Costa et al. 2007, Hase et al. 2008). In an assortative network, $<K_{nn}(k)>$ increases as k increases, while $<K_{nn}(k)>$ in a disassortative network follows decreasing functions of k (Fig 1C). If there are no correlations between degrees of nodes at both ends of a link (*e.g.*, $r = 0$), $<K_{nn}(k)>$ is independent from k and is equal to $<k^2>/<k>$.

It has been shown that the yeast PIN is a disassortative network (Maslov & Sneppen 2002). Therefore, in the yeast PIN, interactions between high- and low-degree nodes are favoured, while those between hubs are suppressed. The biological significance of this structure is unclear. Maslov and Sneppen (2002) proposed that, in the yeast PIN, a hub protein forms a functional module of a cell together with a large number of low-degree neighbors. They then hypothesized that the suppression of links between hubs minimizes unfavourable cross-talks among different functional modules and makes networks robust against perturbations.

If this hypothesis is true, disassortative structure observed in the yeast PIN is under the natural selection, and the disassortativity should be commonly found among PINs in any

organisms. However, by examining PINs from five eukaryote species, Hase et al. (2010) found that the disassortative structure is not a common feature of PINs. The distribution of $<K_{nn}(k)>$ in the PIN can be approximated by $<K_{nn}(k)> \sim k^{-\nu}$, and the value of ν is used to quantify the extent of disassortative structure of a network. Hase et al. (2010) showed that the yeast, worm, fly, and human PINs are disassortative (ν = 0.47, 0.29, 0.35, and 0.26, respectively), while the malaria parasite PIN is not disassortative (ν = 0.02). This observation indicates that the "selectionist view" by Maslov and Sneppen (2002) is not necessary for explaining the disassortative structure of PINs. In section 4, we will see the evolutionary mechanisms generating the difference in assortativity among species.

4. Evolutionary mechanisms generating structures of PINs

To account for the emergence of PIN architecture mentioned above, researchers developed several network growth models and conducted simulation studies using these models. Moreover, statistical properties of PINs were analyzed from the viewpoint of comparative genomics and molecular evolution. In this section, we review evolutionary studies of PINs.

4.1 Preferential attachment and gene duplication

Barabasi and Albert (1999) suggested that the emergence of scale-freeness can be explained by two basic mechanisms: network growth and preferential attachment. The process of network growth adds a new node into a network (red node in Fig 2A). The process of preferential attachment introduces a new link between the new node and each of the other nodes with the probability proportional to the degree of the latter node. For example, the probability that the red node in Fig. 2A gains a new link connected to a blue node is three times higher than that to a black node (Fig 2A). Due to these two processes, a node with a higher degree gains a larger number of links, and thus the degrees of high-degree nodes increase faster than those of low-degree nodes, generating a scale-free network.

In fact, Eisenberg and Levanon (2003) demonstrated that the number of interactions that a protein gained during its evolution is roughly proportional to the degree of the protein by comparing the genomes of *E. coli*, *A. thaliana*, *Schizosaccharomyces pombe*, and *S. cerevisiae*. This observation is consistent with the preferential attachment.

What is the genetic mechanism of network growth and preferential attachment in the evolution of PINs? A plausible mechanism is gene duplication. Let us consider a small PIN containing both high- (node A) and low-degree nodes (node B, C, and D) (Fig 2B, middle). We assume that the number of nodes in a network increases by gene duplication, and a new node has the same interacting partners as the original node. When node B is duplicated, for example, node A acquires a new link and thus the degree of node A increases by one. When node C or node D is duplicated, the same thing happens. On the other hand, if node A is duplicated, each of the degrees of nodes B, C, and D increases by one. Under the assumption that gene duplication occurs randomly with an equal probability for all nodes, the probability that node A acquires a new link is three times higher than the other node does. In general, when we compare a high-degree node (*e.g.*, A) and a low-degree node (*e.g.*, B), a given node (*e.g*, C) is more likely to be a neighbor of a high-degree node than that of a low-degree node. Therefore, a high-degree node gains new links faster than a low-degree node does. For this reason, gene duplication can account for the mechanism of "rich-get-richer".

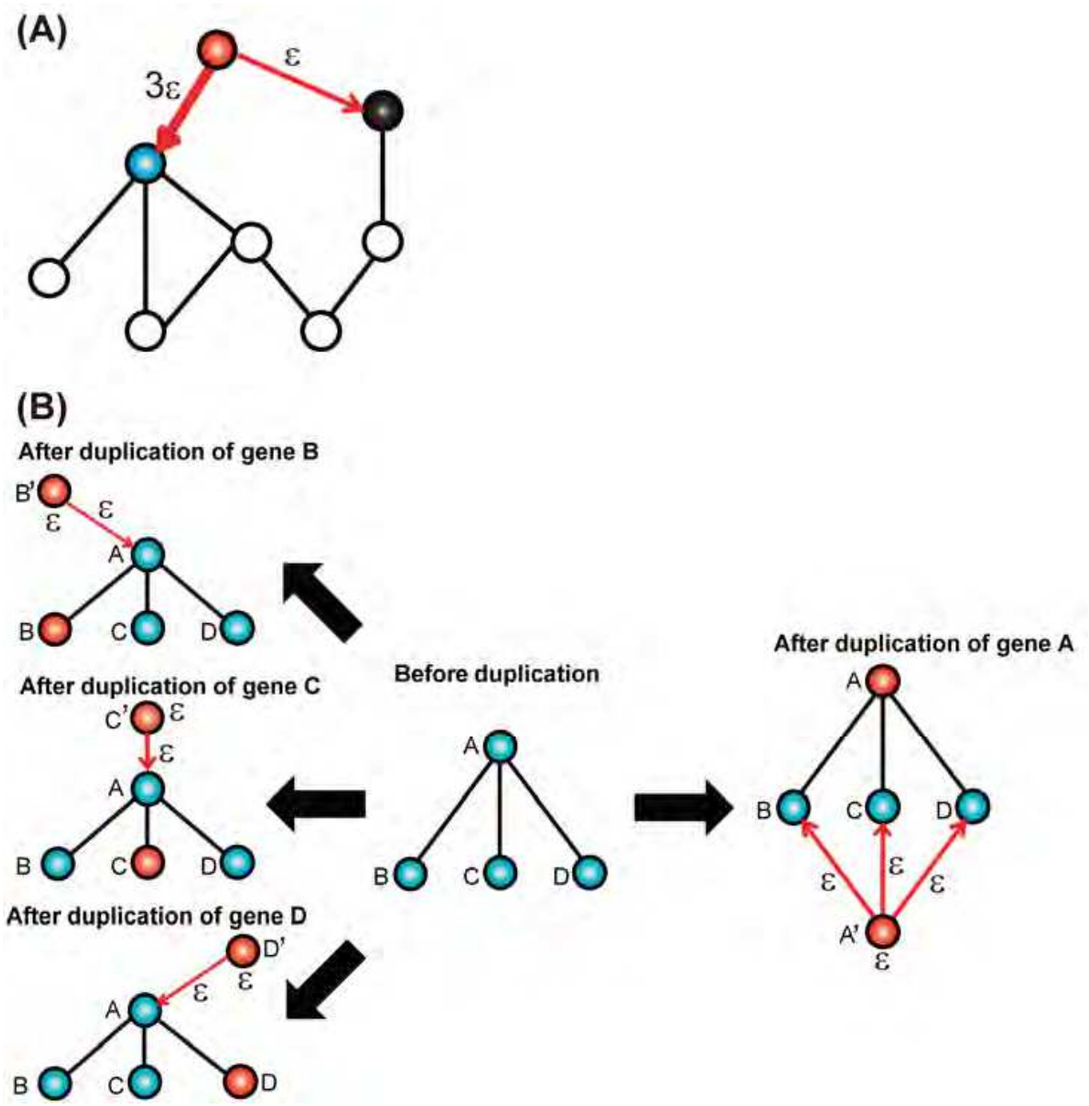


Fig. 2. Network growth by preferential attachment
(A) Preferential attachment. A red node is added to the network. The probability that a new link is attached between red and blue nodes (3ε) is three times higher than that between red and black nodes (ε). **(B)** Network growth with gene duplication. Red nodes represent duplicated nodes. Gene duplication occurs with an equal probability (ε) for all nodes. When node A is duplicated, degrees of nodes B, C, and D increase by one (right), whereas when either node B, C, or D is duplicated, degree of node A increases by one (left).

4.2 Duplication and divergence model

A pair of genes generated by duplication will undergo one of three fates, namely, (i) neofunctionalization, (ii) subfunctionalization, and (iii) nonfunctionalization. After gene duplication, one of the duplicated genes becomes free from selective pressure because of the presence of redundant copies of the gene. Therefore, the gene can tolerate to the accumulation of random mutations and in some cases acquire a novel function (Ohno 1970).

This process is called neofunctionalization. On the other hand, in subfunctionalization process, each of the duplicated genes accumulates mutations, and the functions of the ancestral gene are assigned to the two genes (Force et al. 1999). In nonfunctionalization process, one of the duplicated genes loses its function and becomes a pseudogene due to deleterious mutations. Among the three processes, neofunctionalization and subfunctionalization contribute to the evolution of proteins (Lynch et al. 2000, Blanc et al. 2004, He et al. 2005, Freilich et al. 2006).

In the duplication-divergence model, neofunctionalization and subfunctionalization are modelled as attachment of new links and removal of the links generated by gene duplication, respectively. As for subfunctionalization process, there are two different models, the symmetric divergence and asymmetric divergence. In the former, links are eliminated from both of the duplicated nodes, while in the latter, elimination of links occurs only in one of the two nodes generated by duplication (Fig 3A).

Wagner (2002) reported that one of the duplicated proteins retain a significantly larger number of interactions than the other. For this reason, several network growth models adopted the asymmetric divergence model (Kim et al. 2002, Wagner 2003, Chung et al. 2003, Ispolatov et al. 2005c). However, "complete" asymmetric divergence in which links are eliminated from only one of the duplicates is unrealistic, and the actual divergence process should be intermediate between symmetric and asymmetric divergence (Hase et al. 2010).

Sole et al. (2002) proposed a model on the basis of neofunctionalization and asymmetric divergence. According to their model, after duplication generates a new node, neofunctionalization process attaches a new link between either of the duplicated nodes and each of the other nodes with a uniform probability θ, and then asymmetric divergence eliminates links to only one of the duplicated nodes with a uniform probability α (Fig 3A). Simulation and analytical studies have demonstrated that this model can generate scale-free networks with a small-world property (Sole et al. 2002, Kim et al. 2002, Pastor-Satorras et al. 2003, Chung et al. 2003, Raval 2003).

However, it has been pointed out that some statistical features of PINs could not be regenerated by the model of Sole et al. (2002). The yeast and fly PINs show a much larger $<C>$ than the networks by Sole et al. with the same number of nodes and links as the actual PINs (Sole et al. 2002, Middendorf et al. 2005, Ispolatov et al. 2005a). To overcome this problem, Vazquez et al. (2003) proposed the heterodimerization (HD) model. In their model, symmetric divergence eliminates links from both of the duplicated nodes with a uniform probability α, and the HD process attaches a new link between two duplicated nodes with another uniform probability β, forming a heterodimer (Fig 3A).

When gene duplication occurs for a self-interacting protein, the duplicated proteins will interact to each other. Therefore, β in Vazquez et al. (2003) represents the probability that a randomly selected protein is self-interacting and the new HD link between two duplicated proteins survives after divergence. Simulation and analytical studies have showed that the HD model could reproduce scale-free networks with a similar $<C>$ to the yeast and fly PINs (Vazquez et al. 2003, Middendorf et al. 2005, Ispolatov et al. 2005a). This is because an HD process creates triangles, and a network containing a large number of triangles shows a high value of $<C>$. A computational study based on machine learning technique showed that the

HD model could best reproduce the fly PIN among seven network growth models (Middendorf et al. 2005).

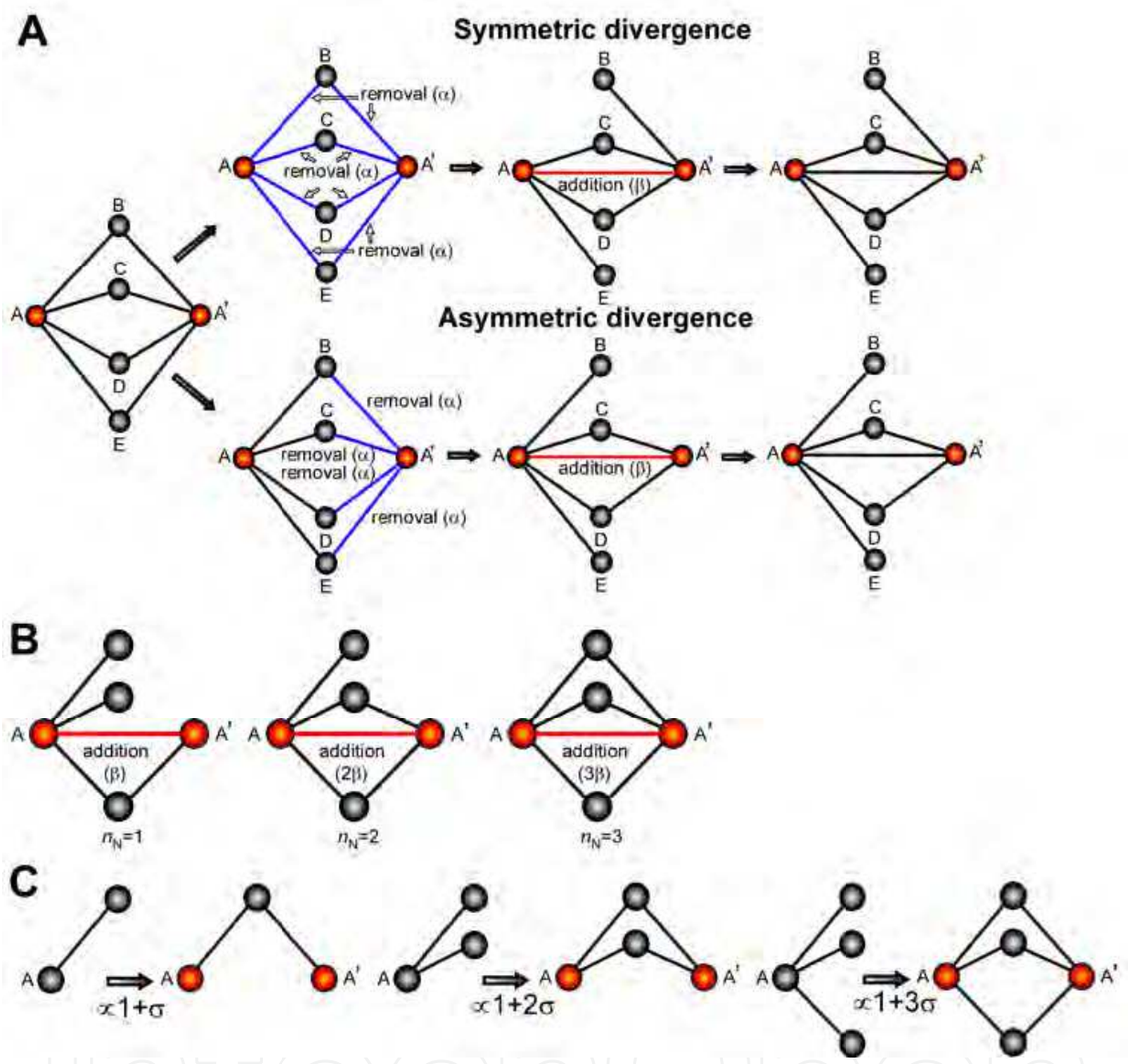


Fig. 3. Network growth models based on gene duplication and divergence
A pair of two red nodes are generated by gene duplication. **(A)** The HD model with asymmetric or symmetric divergence processes. Nodes A and A′ are generated by gene duplication. In the symmetric divergence, each of the links to nodes A and A′ is eliminated with a uniform probability α. On the other hand, in the asymmetric divergence, each of the links to node A′ is eliminated with a uniform probability α. After the divergence process, an HD link (a red line) between two duplicated nodes (nodes A and A′) is attached with a uniform probability β. **(B)** The NHD model. An HD link (red link) is attached between nodes A and A′ with a probability proportional to the number (n_N) of common neighbors shared by these nodes. **(C)** The DDD model. A probability of duplication of a given node is dependent on the degree of the node. If a node has k links, the node is duplicated with the probability proportional to $1 + k\sigma$, where σ is a parameter of the duplicability of a node.

4.3 Non-uniform heterodimerization model

By conducting simulation studies, Hase et al. (2008) showed that, to reproduce the value of <*C*> in the yeast PIN by the HD model, the number of HD links in the networks by the HD model has to be much larger than that in the yeast PIN. Similar observation was made for the fly PIN (Ispolatov et al. 2005a and b). This means that the HD model is insufficient for explaining the evolution of PINs.

As shown in Fig. 3B, when two duplicated nodes share one, two, and three common neighbors, an HD link between them generates one, two, and three new triangles, respectively. The high <*C*> in a PIN indicates that the network contains many triangles. Therefore, if a new HD link is attached more preferentially between duplicated nodes sharing a larger number of common neighbors, the value of <*C*> in a simulation-generated network is expected to become higher. By considering in this way, Hase et al. (2008) proposed the non-uniform heterodimerization (NHD) model in which a new HD link is added between duplicated nodes with a probability proportional to the number of neighbors shared by those nodes (Fig 3B). Simulation studies demonstrated that the NHD model could indeed reproduce both the high value of <*C*> and the small number of HD links in the yeast PIN.

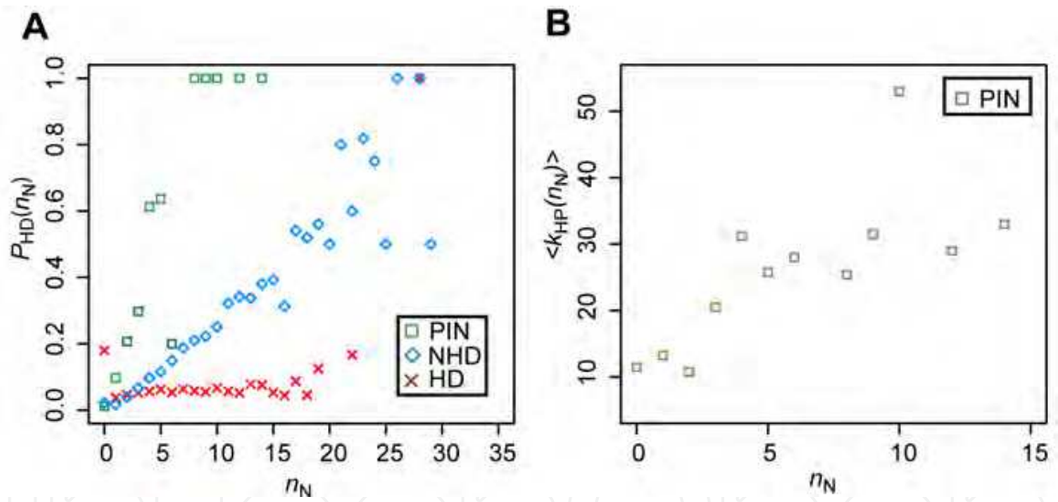


Fig. 4. HD links in the yeast PIN and in the networks by the HD and NHD models (Hase et al. 2008). **(A)** Distribution of $P_{HD}(n_N)$, the probability that an HD link exists between two homologous proteins when they share n_N common neighbors. Green squares, blue diamonds, and red crosses indicate the values for the yeast PIN, the network by the NHD model, and that by the HD model, respectively. **(B)** Distribution of $k_{HP}(n_N)$, the mean degree of proteins that are connected by an HD link and share n_N common neighbors.

In the evolution of PINs, duplication of a self-interacting protein adds an HD link between duplicated proteins. Some HD links were conserved in evolution, while others were eliminated because of occurrence of mutations at interacting sites in these duplicated proteins. In the HD model, the survival rate of HD links is uniform; on the other hand, the NHD model assumes it to be proportional to the number of their common neighbors (Fig. 4A). In the yeast PIN, the probability that two homologous node retain an HD link increases

as the number of neighbors shared by the two nodes increases, which is consistent to the NHD model rather than the HD model (Fig. 4A).

A possible explanation for this observation is as follows. It is expected that, when a given pair of proteins share a large number of common neighbors, the degree of these proteins should be high. In fact, in the yeast PIN, when two homologous proteins are connected by a HD link, there is a positive correlation between the number of common neighbors to the homologues and the mean degree of the two proteins (Fig. 4B). Moreover, several studies showed that high-degree proteins tend to show low evolutionary rate in the yeast PIN (Fraser et al. 2002, 2003, Fraser 2005). Therefore, it is suggested that the survival rates of HD links are also positively correlated with the number of common neighbors shared by the two homologous proteins.

4.4 Degree-dependent duplicability and assortativity

Duplication and divergence models including the NHD model can explain various aspects of the architecture of PINs (Pastor-Satrras et al. 2003, Vazquez 2003, Hase et al. 2008). However, these models cannot explain the differences in overall structures of PINs among species. As mentioned in section 3, the yeast, worm, fly, and human PINs are disassortative, while the malaria parasite PIN is non-disassortative.

A possible evolutionary scenario that can explain the difference in assortativity of PINs among different species is as follows (Hase et al. 2010). Let us consider a disassortative network containing low- and high-degree nodes (*e.g.*, A and C, respectively), in which the low- and high-degree nodes are linked to each other (Fig 5A, middle). Duplication of a low-degree node (*e.g.*, node A) causes the value of ν in the disassortative network to be higher, because the degree of its high-degree neighbor increases (Fig 5A, left). On the other hand, duplication of a high-degree node (*e.g.*, node C) makes the degree of its low-degree neighbors higher, and thus the value of ν decreases (Fig 5A, right). For this reason, duplication of low- and high-degree nodes would make the value of ν in a disassortative network larger and smaller, respectively.

Hase et al. (2010) proposed a novel duplication and divergence model named "degree–dependent duplication (DDD) model", in which duplication of nodes occurs depending on their degree (see Fig 3C). Simulation studies based on the DDD model revealed that preferential duplication of low-degree nodes can successfully reproduce the disassortative structure observed in the yeast, worm, and fly PINs, while preferential duplication of high-degree nodes generate non-disassortative networks similar to the malaria parasite PIN (see Fig 5B and 5C). Moreover, Hase et al. (2010) evaluated the dependency of gene duplicability on their degrees by analyzing orthologous relationships of genes extracted from 55 eukaryotic proteomes. The analyses demonstrated that proteins with a lower degree indeed have higher duplicability in disassortative PINs (the yeast, worm, and fly PINs) (Fig 5D), whereas high-degree proteins tend to have high duplicability in non-disassortative PINs (the malaria parasite PIN) (Fig 5E). Therefore, it is suggested that assortativity of a PIN is related with the gene duplicability dependent on the degrees of genes. If this is the case, disassortative structure of PINs is merely a byproduct of preferential duplication of low-degree proteins, and we do not need to assume any adaptive meaning for this structure, as mentioned in section 3.

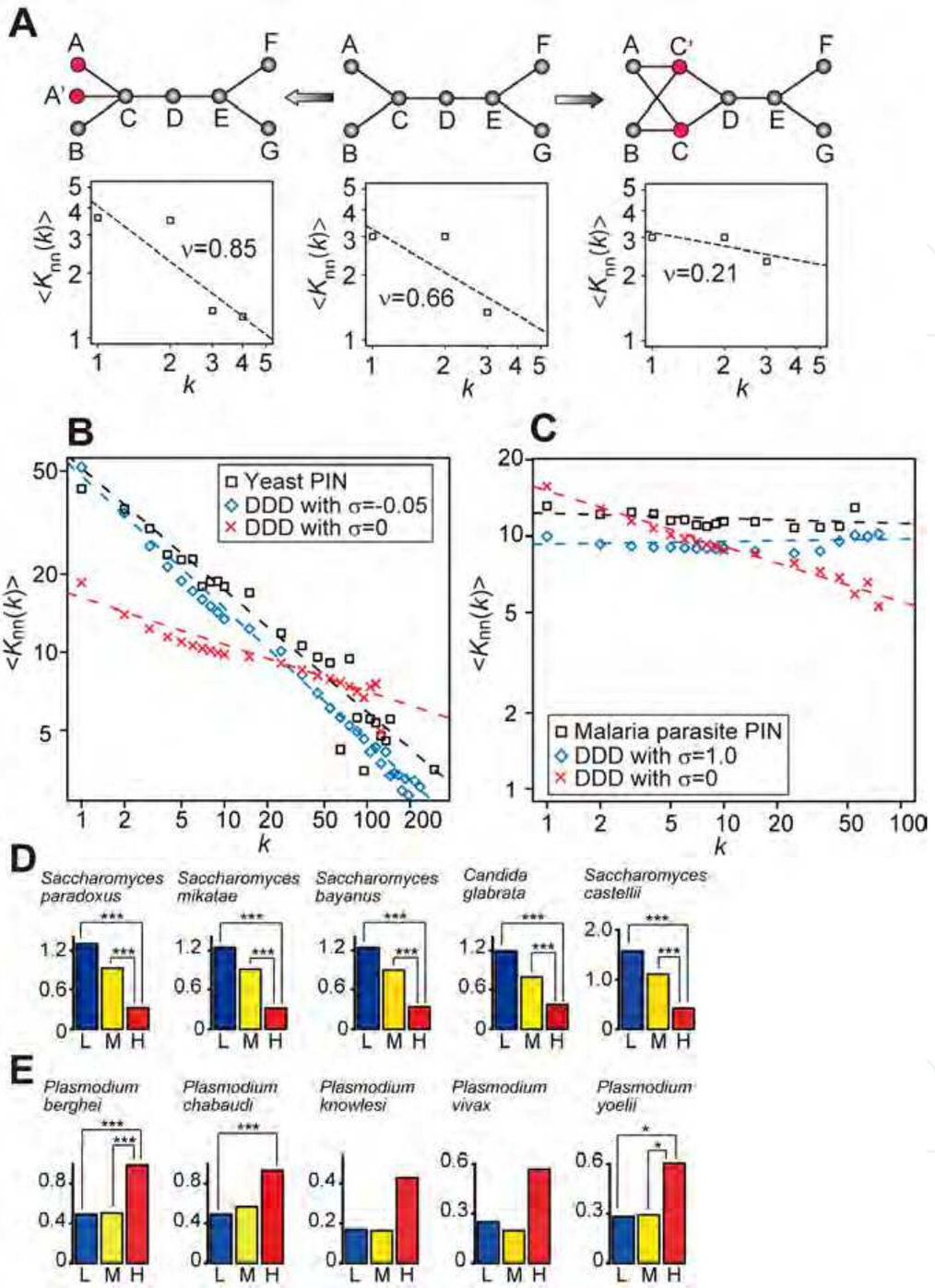


Fig. 5. The DDD model and the extent of assortativity in networks (Hase et al. 2010).
(A) Duplication of a node alters the distribution of $<K_{nn}(k)>$ and the value of ν in a network. A diagram below a network shows the distribution of $<K_{nn}(k)>$ and the value of ν in the network. **(B)** The distribution of $<K_{nn}(k)>$ in the networks generated by the DDD model for yeast. Blue diamonds and red crosses show the results of simulation with σ = -0.05 and 0, respectively (as for σ, see Fig. 3C). Black squares represent $<K_{nn}(k)>$ in the yeast PIN. Dashed

lines in black, blue, and red represent $k^{-0.47}$, $k^{-0.51}$, and $k^{-0.18}$, respectively. **(C)** The distribution of $<K_{nn}(k)>$ in the networks generated by the DDD model for malaria parasite. Blue diamonds and red crosses show the results of simulation with $\sigma = 1.0$ and 0, respectively. Black squares represent $<K_{nn}(k)>$ in the malaria parasite PIN. Dashed lines in black, blue, and red represent $k^{-0.02}$, $k^{0.01}$, and $k^{-0.22}$, respectively. **(D)** and **(E)** indicate correlations between the degree and the duplicability in the yeast and malaria parasite PINs, respectively. Bars in blue, yellow, and red show the mean duplicability among low-, middle-, and high-degree proteins, respectively. A species name above each diagram denotes the species of which genome was compared with *S. cerevisiae* or *P. falciparum*. *, **, and *** represent $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively, by the Wilcoxon rank-sum test with the Bonferroni correction.

5. Structures of PINs and their relationships with disease genes and drug targets

As we have seen above, PINs are characterized by several interesting properties that are different from those of a random network. Therefore, understanding diseases and mechanisms of drug action in the context of PIN architecture may allow us to address some fundamental properties of disease genes and drug target molecules. Indeed, number of disease genes and that of drug targets are very small. Only 10% of the human genes are known to be disease genes (Amberger et al. 2009), and only 435 genes are target genes of therapeutic drugs (Rask-Andersen et al. 2011). Why is the number of drug targets and disease genes so small? Are they distributed randomly over the human PIN? Are there any quantifiable correlations between drug target genes and their statistical properties in the human PIN? To address these questions, drug target and disease genes were mapped onto the human PIN and their statistical properties in the PIN were investigated. Moreover, by using biological networks including the human PIN data, several studies showed that side effects of drugs depend on their statistical features in the network. In this and subsequent sections, we review the application of network analyses to medical researches.

5.1 Statistical properties of disease genes and drug targets in the human PIN

Elimination of a hub protein affects many proteins in a network (Jeong et al. 2001, Yu et al. 2008). Therefore, it was previously hypothesized that genes encoding hub proteins are associated with diseases (Barabasi et al. 2011). Several studies reported that the mean degree among disease genes is in fact significantly higher than that among non-disease genes (Wachi et al. 2005, Jonsson & Bates 2006, Xu & Li 2006).

A human gene is defined to be essential, when knock-out of its orthologous gene causes embryonic and postnatal lethality or sterility in mouse (Liang & Li 2007). Liang & Li (2007) reported that essential genes tend to encode hub proteins in the human PIN.

However, Wachi et al. (2005), Jonsson & Bates (2006), and Xu & Li (2006) took no account for the fact that there are only a small number of disease genes that are also essential (essential disease genes), while vast majority of disease genes are actually non-essential. Because essential disease genes encode hub proteins, the mean degree of disease genes became apparently high in the three studies. In contrast, non-essential disease genes do not show any tendency to encode hub proteins (Goh et al. 2007). Rather, they tend to encode low- and

middle-degree proteins (Feldman et al. 2008). Mutations in high-degree proteins cause dysfunctionality of many neighbor proteins, leading severe impairment of developmental and physiological processes. Individuals having such mutations cannot survive until reproductive years and are likely to be removed from population. For this reason, non-essential disease genes are enriched among low- and middle-degree genes.

Hase et al. (2009) investigated drug target genes to see whether they have specific statistical features in the PIN or not. They found that most drug target genes are middle-degree proteins and some are low-degree, while there are almost no drug targets among high-degree proteins (see Fig 6). The degree distribution is similar to that of disease genes, and, not surprisingly, drug target genes significantly overlap with disease genes (Yao & Rzhetsky 2008). These results indicate that middle-degree proteins are likely to be most advantageous targets for therapeutic drugs.

Oncogenes tend to be high-degree proteins (Jonsson & Bates 2006), and thus they are less likely to be targets for drugs, or one must accept major potential side effects. A possible strategy for designing anti-cancer therapy with less severe side effects is to develop a novel combination of drug compounds that targets several low- and middle-degree proteins, because such combination could generate synergetic effects to cancer cure, and low- or middle-degree targets are expected to induce less severe side effects compared with high-degree targets.

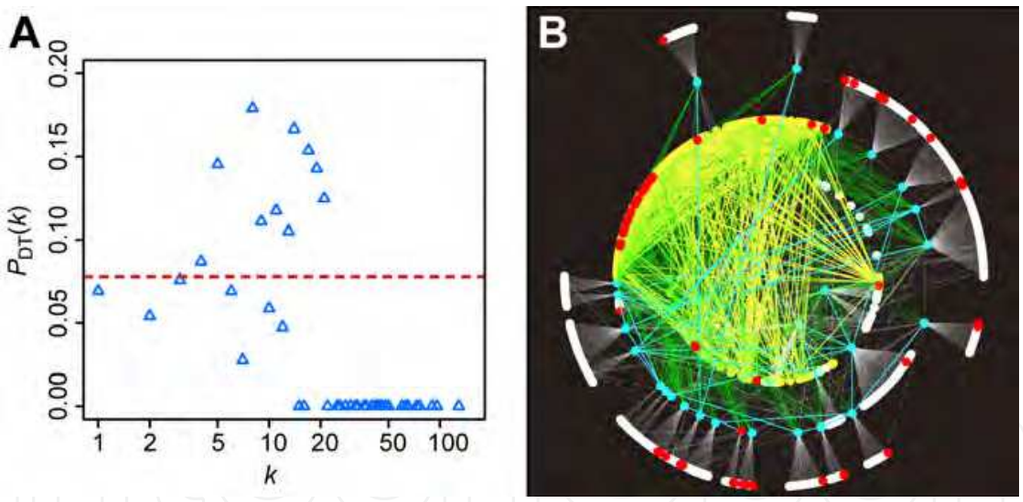


Fig. 6. Degree distribution of drug targets (Hase et al. 2009).
(A) $P_{DT}(k)$ represents the fraction of drug targets to all proteins for the degree of k. The dashed line in red represents the probability that a randomly selected protein is a drug target. **(B)** White, yellow, and blue nodes represent low- (k = 1 – 5), middle- (k = 6 – 30) and high-degree (k > 30) proteins, respectively. Drug targets (red nodes) are mapped on the network. White, yellow, green, and blue links represent interactions between high- and low-degree proteins, those between middle-degree proteins, those between high- and middle-degree proteins, and those between high-degree proteins, respectively. Middle-degree proteins are extensively connected to each other, while links between high-degree proteins are rather suppressed. For clarity, low- and middle-degree proteins that do not have any interactions with high-degree proteins are not shown.

5.2 Predicting candidate drug targets and their side effects based on biological networks

To develop a new drug, it is critical to accurately predict its side effect, because almost 30% of candidate drugs are rejected in clinical stages due to their unexpected toxicity or concerns about drug safety (Kola & Landis 2004, Billingsley 2008). Severe adverse reactions may be found long after the approval of drugs (*e.g.*, Rosiglitazone), and in such cases, those drugs would go out of production (*e.g.*, Rofecoxib) (Moore et al. 2007).

The chemical structures of drugs have been used to predict their adverse side effects and target proteins (Kuhn et al. 2008, Campillos et al. 2008, Yamanishi et al. 2010). Campillos et al. (2008) developed a large-scale database of adverse side effects of drugs. By using the database with information of chemical structure of drugs, they made a similarity metric between two drugs. Under the assumption that drugs with higher similarity in their metric more tend to share the same target proteins, they inferred candidate targets for the drugs.

However, if target proteins of two drugs are close in a molecular network, such drugs may cause similar downstream effects in the network and thus have similar side effects. To understand the molecular mechanisms of drug action and associated adverse effects in greater details, it makes sense to view targets of drugs in the context of biological networks including the genome-wide human interactome (Pache et al. 2008, Zanzoni et al. 2009).

Recently, Brouwers et al. (2011) investigated how side effect similarities of targets depend on their closeness in the human PIN. They found that a certain number of pairs of two drugs without common targets show similar side effects, when they are close in the human PIN. Moreover, Wang et al. (2011) reported that drug side effects are significantly associated with network distances between drug target genes and diseases genes, *i.e.*, targets for failure drugs that make severe adverse side effects are closer to disease genes than targets for approved drugs. Thus, selecting targets that are too close to diseases genes are not always the best strategy (Wang et al. 2011), although the pharmaceutical industry tends to select targets of new drugs that are close with the corresponding disease genes in the biological networks, especially after 1996 (Yildirim et al. 2007).

With recent influx of information of biological networks, especially the human interactome, analyses like Brouwers et al. (2011) or Wang et al. (2011) can be refined and adapted to infer still unknown adverse side effects of drugs and to predict possible target genes. Indeed, by integrating information of the human PIN with similarities between two genes (*e.g.*, GO semantic and sequence similarity) and those between two drugs (*e.g.*, chemical and drug therapeutic similarity), several recent researches attempted to develop a method to predict possible targets for therapeutic drugs (Zhao & Li 2010, Perlman et al. 2011).

6. Possibilities of medical studies with integration of PINs and evolutionary studies

The human PIN is still incomplete and there are many proteins without any information of protein-protein interactions (Venkatesan et al. 2009). Evolutionary information (*e.g.*, evolutionary rate and duplicability) of genes is significantly correlated with their statistical properties in PINs (see sections 2 and 3); therefore, such information can be utilized to complement to the incompleteness of the human PIN.

Rambaldi et al. (2008) reported that most of the cancer genes are singletons and have interactions with many genes. This finding indicates that both gene duplicability and network information are useful for predicting candidate cancer genes. Modification of currently available methods by integrating evolutionary information would improve the accuracy of predicting disease and drug target genes.

Recently, large-scale PINs became available from several prokaryotes, including Methicillin-resistant *Staphylococcus aureus* (MRSA) (Cherkasov et al. 2010), *Treponema pallidum* (Titz et al. 2008), *Campylobacter jejuni* (Parrish et al. 2007), *Mycoplasma pneumonia* (Kuhner et al. 2009), and *Mesorhizobium loti* (Shimoda et al. 2008) (see Table 1). Some of them are pathogenic. By investigating the evolution of their PINs, we may be able to understand the process of acquiring the pathogenicity and developing drug resistance from the viewpoint of network architecture.

Cherkasov et al. (2010) suggested that, in the MRSA PIN, hubs are essential for network stability and may be prospective antimicrobial drug targets. However, almost all known antimicrobial targets have relatively few interactions and hubs have largely been overlooked as drug targets. If hubs in pathogens have no orthologous genes in human and evolve very slowly, by targeting such hubs, we may be able to develop novel antibacterial drugs with high efficacy and small side effects, and without development of resistance to the drugs. With a recent influx of PINs from pathogenic organisms and genomes from various bacterial species, analyses integrating comparative genomics with PINs will become keys to identify still unknown disease mechanisms and novel targets for antibacterial drugs.

7. Conclusion

In this chapter, we describe various aspects of architecture of PINs, such as scale-freeness, small-world properties, and assortativity. Computational studies based on network-growth models and comparative genomics revealed how accumulation of local changes in PINs affects their overall architecture during evolution. We also discussed possible application of PINs and evolutionary studies to medical researches. With expected explosion of OMICs data (*e.g.*, PINs and SNPs from human) in the near future, an integration of networks and genetics will be among the most powerful strategies to elucidate unknown mechanisms of disorders and discover novel targets for efficacious drugs.

8. Acknowledgements

This study was supported by the Ministry of Education, Culture, Sports, Science and Technology, Japan, grant 23770271 to YN.

9. References

Albert, R.; Jeong, H. & Barabasi, AL. (1999). Diameter of the World-Wide Web, *Nature*, vol. 401, pp. 130–131.

Amberger, J.; Bocchini, CA.; Scott, AF. & Harmosh, A. (2009). McKusick's Online Mendelian Inheritance in Man (OMIM), *Nucleic Acids Research*, vol. 37, pp. D793–D796.

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an Arabidopsis interactome map, *Science*, vol. 333, pp. 601–607.

Arifuzzaman, M.; Maeda, M.; Itoh, A.; Nishikata, K.; Takita, C.; Saito, R.; Ara, T.; Nakahigashi, K.; Huang, HC.; Hirai, A.; Tsuzuki, K.; Nakamura, S.; Altaf-Ul-Amin, M.; Oshima, T.; Baba, T.; Yamamoto, N.; Kawamura, T.; Ioka-Nakamichi, T.; Kitagawa, M.; Tomita, M.; Kanaya, S.; Wada, C. & Mori, H. (2007). Large-scale identification of protein-protein interaction of Escherichia coli K-12, *Genome Research*, vol. 16, pp. 686–691.

Barabasi, AL. & Albert, R. (1999). Emergence of scaling in random networks, *Science*, vol. 286, pp. 509–512.

Barabasi, AL. & Oltvai, ZN. (2004). Network biology : understanding the cell's functional organization, *Nature Reviews Genetics*, vol. 5, pp. 101–113.

Barabasi, AL.; Gulbahce, N. & Lascalzo, J. (2011). Network medicine : a network-based approach to human disease, *Nature Reviews Genetics*, vol. 12, pp. 56–68.

Billingsley, ML. (2008). Druggable targets and targeted drugs: enhancing the development of new therapeutics, *Pharmacology* vol. 82, pp. 239–244.

Blanc, G. & Wolfe, KH. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution, *The Plant Cell*, vol. 16, pp. 1679–1691.

Brouwers, L.; Iskar, M.; Zeller, G.; Noort, V. & Bork, P. (2011). Network neighbors of drug targets contribute to drug side effect similarity, *PLoS ONE*, vol. 6, e22187.

Callaway, DS.; Hopcroft, JE.; Kleinberg, JM.; Newman, MEJ. & Strogatz, SH. (2001). Are randomly grown graphs really random?, *Physical Review E*, vol. 64, 041902.

Campillos, M.; Kuhn, M.; Gavin, AC.; Jensen, LJ. & Bork, P. (2008). Drug target identification using side effect similarity, *Science*, vol. 321, pp. 263–266.

Cherkasov, A.; Hsing, M.; Zoraghi, R.; Foster, LJ.; See, RH.; Stoynov, N.; Jiang, J.; Kaur, S.; Lian, T.; Jackson, L.; Gong, H.; Swayze, R.; Amandoron, E.; Hormozdiari, F.; Dao, P.; Sahinalp, C.; Santos-Filho, O.; Axerio-Cilies, P.; Byler, K.; McMaster, WR.; Brunham, RC.; Finlay, BB. & Reiner, NE. (2011). Mapping the protein interaction network in Methicillin-resistant Staphylococcus aureus, *Journal of Proteome Research*, vol. 10, pp. 1139–1150.

Chung, F. & Lu, L. (2002). The average distances in random graphs with given expected degrees, *Proceedings of National Academy of Sciences USA*, vol. 99, pp. 15879–15882.

Chung, F.; Lu, L.; Dewey, TG. & Galas DJ. (2003). Duplication models for biological networks, *Journal of computational biology*, vol. 10, pp. 677–687.

Cohen, R. & Havlin, S. (2003). Scale-free networks are ultrasmall, *Physical Review Letters*, vol. 90, 058701.

Costa, LF.; Rodrigues, FA.; Travieso, G. & Boas, RRV. (2007). Characterization of complex networks : A survey of mesurements, *Advances in Physics*, vol. 56, pp. 167–242.

Deane CM.; Salwinski L.; Xenarios I. & Eisenberg D. (2002) Protein interactions, *Molecular & Cellular Proteomics*, vol. 1, pp. 349–356.

Eisenberg, E. & Levanon, EY. (2003). Preferential attachment in the protein network evolution, *Physical Review Letters*, vol. 91, 138701.

Erdös, P. & Renyi, A. (1960). On the evolution of random graphs, *Publication of the Mathematical Institute of the Hungarian Academy of Science*, vol. 5, pp. 17–61.

Feldman, I.; Rzhetsky, A. & Vitkup, D. (2008). Network properties of genes harbouring inherited disease mutations, *Proceedings of National Academy of Sciences USA*, vol. 105, pp. 4323–4328.

Force, A.; Lynch, M.; Pickett, FB.; Amores, A.; Yan, YL. & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations, *Genetics*, vol. 151, pp. 1531–1545.

Fraser, HB.; Hirsh, AE.; Steinmetz, LM.; Scharfe, C. & Feldman, MW. (2002). Evolutionary rate in the protein interaction network, *Science*, vol. 296, pp. 750–752.

Fraser, HB.; Wall, DP. & Hirsh, AE. (2003). A simple dependence between protein evolution rate and the number of protein-protein interactions, *BMC Evolutionary Biology*, vol. 3, 11.

Fraser, HB. (2005). Modularity and evolutionary constraint on proteins, *Nature Genetics*, vol. 37, pp. 351–352.

Freilich, S.; Massingham, T.; Blanc, E.; Goldovsky, L. & Thornton, JM. (2006). Relating tissue specialization to the differentiation of expression of singleton and duplicate mouse protein, *Genome Biology*, vol. 7, R89.

Giot, L.; Bader, JS.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, YL.; Ooi, CE.; 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, CA.; Finley, RL Jr.; White, KP.; Braverman, M.; Jarvie, T.; Gold, S.; Leach, M.; Knight, J.; Shimkets, RA.; McKenna, MP.; Chant, J. & Rothberg, JM. (2003). A protein interaction map of Drosophila melanogaster, *Science*, vol. 302, pp. 1727–1736.

Goh, KI.; Cusick, ME.; Valle, D.; Childs, B.; Vidal, M. & Barabasi, AL. (2007). The human disease network, *Proceedings of National Academy of Sciences USA*, vol. 104, pp. 8685–8690.

Guldener, U.; Munsterkotter, M.; Oesterheld, M.; Pagel, P.; Ruepp, A.; Mewes, HW. & Stumpflen, V. (2006). Mpact: the MIPS protein interaction resource on yeast, *Nucleic Acids Research*, vol. 34, pp. D436–441.

Hakes, L.; Pinney, JW.; Robertson, DL. & Lovell, SC. (2008). Protein-protein interaction networks and biology–what's the connection?, *Nature Biotechnology*, vol. 26, pp. 69–72.

Han, JDJ.; Bertin, N.; Hao, T.; Goldberg, DS.; Berriz, GF.; Zhang, LV.; Dupuy, D.; Walhout, AJM.; Cusick, ME.; Roth, FP. & Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network, *Nature*, vol. 430, pp. 88–93.

Hase, T.; Niimura, Y.; Kaminuma, T. & Tanaka, H. (2008). Non-uniform survival rate of heterodimerization links in the evolution of the yeast protein-protein interaction network, *PLoS ONE*, vol. 3, e1667.

Hase, T.; Tanaka, H.; Suzuki, Y.; Nakagawa, S. & Kitano, H. (2009). Structure of protein interaction networks and their implications on drug design, *PLoS Computational Biology*, vol. 5, e1000550.

Hase, T.; Niimura, Y. & Tanaka, H. (2010). Difference in gene duplicability may explain the difference in overall structure of protein-protein interaction networks among eukaryotes, *BMC Evolutionary Biology*, vol. 10, 358.

He, X. & Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution, *Genetics*, vol. 169, pp. 1157–1164.

Ispolatov, I.; Krapivsky, PL.; Mazo, I. & Yuryev, A. (2005a). Cliques and duplication-divergence network growth, *New Journal of Physics*, vol. 7, 145.

Ispolatov, I.; Yuryev, A.; Mazo, I. & Maslov, S. (2005b). Binding properties and evolution of homodimers in protein-protein interaction networks, *Nucleic Acids Research*, vol. 33, pp. 3629–3635.

Ispolatov, I.; Krapivsky, PL. & Yuryev, R. (2005c). Duplication-divergence model of protein interaction network, *Physical Review E*, vol. 71, 061911.

Ito, T.; Chiba, T.; Ozawa, R.; Yoshida, M.; Hattori, M. & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proceeding of National Academy of Sciences USA*, vol. 98, pp. 4569–4574.

Jeong, H.; Mason, SP.; Barabasi, AL. & Oltvai, ZN. (2001). Lethality and centrality in protein interaction networks, *Nature*, vol. 411, pp. 41–42.

Jonsson, PF. & Bates, PA. (2006). Global topological features of cancer proteins in the human interactome, *Bioinformatics*, vol. 22, pp. 2291–2297.

Kim, J.; Krapivsky, PL.; Kahng, B. & Redner, S. (2002). Infinite-order precolation and giant fluctuations in a protein interaction network, *Physical Review E*, vol. 66, 055101.

Kola, I. & Landis, J. (2004). Can the phamaceutical industry reduce attrition rates?, *Nature Reviews Drug Discovery*, vol. 3, pp. 711–715.

Kuhn, M.; Campillos, M.; Gonzalez, P.; Jensen, LJ. & Bork, P. (2008). Large-scale prediction of drug-target relationships, *FEBS Letters*, vol. 582, pp. 1283–1290.

Kuhner, S.; Noort, VV.; Betts, MJ.; Leo-Macias, A.; Batisse, C.; Rode, M.; Yamada, T.; Maier, T.; Bader, S.; Beltran-Alvarez, P.; Castano-Diez, D.; Chen, WH.; Devos, D.; Guell, M.; Norambuena, T.; Racke, I.; Rybin, V.; Schmidt, A.; Yus, E.; Aebersold, R.; Herrmann, R.; Bottcher, B.; Frangakis, AS.; Russell, RB.; Serrano, L.; Bork, P. & Gavin, AC. (2009). Proteome organization in a genome-reduced bacterium, *Science*, vol. 326, pp. 1235–1240.

LaCount, DJ.; Vignali, M.; Chettier, R.; Phansalkar, A.; Bell, R.; Hesselberth, JR.; Schoenfeld, LW.; Ota, I.; Sahasrabudhe, S.; Kurschner, C.; Fields, S. & Hughes, RE. (2005). A protein interaction network of the malaria parasite Plasmodium falciparum, *Nature*, vol. 438, pp. 103–107.

Lavallee-Adam, M.; Cloutier, P.; Coulombe, B. & Blanchette, M. (2011) Modeling contaminants in AP-MS/MS experiments, *Journal of Proteome Research*, vol. 10, pp. 886–895.

Li, S.; Armstrong, CM.; Bertin, N.; Ge, H.; Milstein, S.; Boxem, M.; Vidalain, PO.; Han, JD.; Chesneau, A.; Hao, T.; Goldberg, DS.; Li, N.; Martinez, M.; Rual, JF.; Lamesch, P.; Xu, L.; Tewari, M.; Wong, SL.; Zhang, LV.; Berriz, GF.; Jacotot, L.; Vaglio, P.; Reboul, J.; Hirozane-Kishikawa, T.; Li, Q.; Gabel, HW.; Elewa, A.; Baumgartner, B.; Rose, DJ.; Yu, H.; Bosak, S.; Sequerra, R.; Fraser, A.; Mango, SE.; Saxton, WM.; Strome, S.; Heuvel, VDS.; Piano, F.; Vandenhaute, J.; Sardet, C.; Gerstein, M.; Doucette-Stamm, L.; Gunsalus, KC.; Harper, JW.; Cusick, ME.; Roth, FP.; Hill, DE. & Vidal, M. (2004). A map of the interactome network of the metazoan C. elegans, *Science*, vol. 303, pp. 540–543.

Liang, H. & Li, WH. (2007). Gene essentiality, gene duplicability and protein connectivity in human and mouse, *Trends in Genetics*, vol. 23, pp. 375–378.

Lynch, M. & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization, *Genetics*, vol. 154, pp. 459–473.

Maslov, S. & Sneppen, K. (2002). Specificity and stability in topology of protein networks, *Science*, vol. 296, pp. 910–913.

Middendorf, M.; Ziv, E. & Wiggins, CH. (2005). Inferring network mechanisms : the Drosophila melanogaster protein interaction network, *Proceedings of National Academy of Sciences USA*, vol. 102, pp. 3192–3197.

Moore, TJ.; Cohen, MR. & Furberg, CD. (2007). Serious adverse drug events reported to Food and Drug Administration, 1998-2005, *Archives of Internal Medicine*, vol. 167, pp. 1752–1759.

Newman, MEJ. (2002). Assortative mixing in networks, *Phyical Review Letters*, vol. 89, 208701.

Ohno, S. (1970). *Evolution by gene duplication*, Springer-Verlag, New-York, USA.

Pache, RA.; Zanozoni, A.; Naval, J.; Mas, JM. & Aloy, P. (2008). Towards a molecular characterisation of pathological pathways, *FEBS Letters*, vol. 582, pp. 1259–1265.

Pacifico, S.; Liu, G.; Guest, S.; Parrish, JR.; Fotouhi, F. & Finley, RL. (2006). A database and tool, IM Browser, for exploring and integrating emerging gene and protein interaction data for Drosophila, *BMC Bininformatics*, vol. 7, 195.

Parrish, JR.; Gulyas, KD. & Finley, RL. (2006) . Yeast two-hybrid contributions to interactome mapping, *Current Opinion in Biotechnology*, vol. 17, pp. 387–393.

Parrish, JR.; Yu, J.; Liu, G.; Hines, JA.; Chan, JE.; Mangiola, BA.; Zhang, H.; Pacifico, S.; Fotouhi, F.; DiRita, VJ.; Ideker, T.; Andrew, P. & Finley, RL. (2007). A proteome-wide protein interaction map for Campylobacter jejuni, *Genome Biology*, vol. 8, R130

Pastor-Satorras, R.; Vazquez, A. & Vespignani, A. (2001). Dynamical and correlation properties of the internet, *Physical Review Letters*, vol. 87, 258701.

Pastor-Satorras, R.; Smith, ED. & Sole, RV. (2003). Evolving protein interaction networks through gene duplication, *Journal of Theorerical Biology*, vol. 222, pp. 199–210.

Perlman, L.; Gottlieb, A.; Atias, N.; Ruppin, E. & Sharan, R. (2011). Combining drug and gene simiality measures for drug-target elucidation, *Journal of Computational Biology*, vol. 18, pp. 133–145.

Rambaldi, D.; Giorgi, FM.; Capuani, F.; Ciliberto, A. & Ciccarelli, FD. (2008). Low duplicability and network fragility of cancer genes, *Trends in Genetics*, vol. 24, pp. 427–430.

Rask-Andersen, M.; Almen, MS. & Schioth, HB. (2011). Trends in the exploitation of novel drug targets, *Nature Reviews Drug Discovery*, vol. 10, pp. 579–590.

Raval, A. (2003). Some asymptotic properties of duplication graphs, *Physical Review E*, vol. 68, 066119.

Reguly, T.; Breitkreutz, A.; Boucher, L.; Breitkreutz, BJ.; Hon, GC.; Myers, CL.; Parsons, A.; Friesen, H.; Oughtred, R.; Tong, A.; Stark, C.; Ho, Y.; Botstein, D.; Andrews, B.; Boone, C.; Troyanskya, OG.; Ideker, T.; Dolinski, K.; Batada, NN. & Tyers, M. (2006). Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae, *Journal of Biology*, vol. 5, 11.

Rual, JF.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, GF.; Gibbons, FD.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem,

M.; Milstein, S.; Rosenberg, J.; Goldberg, DS.; Zhang, LV.; Wong, SL.; Franklin, G.; Li, S.; Albala, JS.; Lim, J.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, RS.; Vandenhaute, J.; Zoghbi, HY.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, ME.; Hill, DE.; Roth, FP. & Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network, *Nature*, vol. 437, pp. 1173–1178.

Shimoda, Y.; Shinpo, S.; Kohara, M.; Nakamura, Y.; Tabata, S. & Sato, S. (2008). A large scale analysis of protein-protein interactions in Nitrogen-fixing Bacterium Mesorhizobium loti, *DNA Research*, vol. 15, pp. 13–23.

Simonis, N.; Rual, JF.; Carvunis, AR.; Tasan, M.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Sahalie, JM.; Venkatesan, K.; Gebreab, F.; Cevik, S.; Klitgord, N.; Fan, C.; Braun, P.; Li, N.; Ayivi-Guedehoussou, N.; Dann, E.; Bertin, N.; Szeto, D.; Dricot, A.; Yildirim, MA.; Lin, C.; Smet, AS.; Kao, HL.; Simon, S.; Smolyar, A.; Ahn, JS.; Tewari, M.; Boxem, M.; Milstein, S.; Yu, H.; Dreze, M.; Vandenhaute, J.; Gunsalus, KC.; Cusick, ME.; Hill, DE.; Tavernier, J.; Roth, FP. & Vidal, M. (2009). Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network, *Nature Methods*, vol. 6, pp. 47–54.

Sole, RV.; Pastor-Satorras, R.; Smith, ED. & Kepler, T. (2002). A model of large-scale proteome evolution, *Advances in Complex Systems*, vol. 5, pp. 43–54.

Stelzl, U.; Worm, U.; Lalowski, M.; Haenig, C.; Brembeck, FH.; 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. & Wanker, EE. (2005). A human protein-protein interaction network: a resource for annotating proteome, *Cell*, vol. 122, pp. 957–968.

Titz, B.; Rajagopala, SV.; Goll, J.; Hauser, R.; McKevitt, MT.; Palzkill, T. & Uetz, P. (2008). The binary protein interactome of Treponema pallidum—the syphilis spirochete, *PLoS ONE*, vol. 3, e2292.

Uetz, P.; Giot, L.; Cagney, G.; Mansfield, TA.; Judson, RS.; Knight, JR.; 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. & Rothberg, JM. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae, *Nature*, vol. 403, pp. 623–627.

Vazquez, A.; Flammini, A.; Maritan, A. & Vespignani, A. (2003). Modeling of protein interaction networks, *Complexus*, vol. 1, pp. 38–44.

Vazquez, A. (2003). Growing networks with local rules : preferential attachment, clustering hierarchy, and degree correlations, *Physical Review E*, vol. 67, 056104.

Venkatesan, K.; Rual, JF.; Vazquez, A.; Stelzl, U.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Zenkner, M.; Xin, X.; Goh, KI.; Yildirim, MA.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, JM.; Cevik, S.; Simon, C.; Smet, AS.; Dann, E.; Smolyar, A.; Vinayagam, A.; Yu, H.; Szeto, D.; Borick, H.; Dricot, A.; Klitgord, N.; Murray, RR.; Lin, C.; Lalowski, M.; Timm, J.; Rau, K.; Boone, C.; Braun, P.; Cusick, ME.; Roth, FP.; Hill, DE.; Tavernier, J.; Wanker, EE.; Barabasi, AL. & Vidal, M. (2009). An empirical framework for binary interactome mapping, *Nature Methods*, vol. 6, pp. 83–90.

Vidal, M.; Cusick, ME. & Barabasi, AL. (2011). Interactome networks and human disease, *Cell*, vol. 144, pp. 986–998.

Wachi, S.; Yoneda, K. & Wu, R. (2005). Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues, *Bioinformatics*, vol. 21, pp. 4205–4208.

Wang, J.; Li, ZX.; Qui, CX.; Wang, D. & Cui, QH. (2011). The relationship between rational drug design and drug side effects, *Briefings in Bioinformatics*, (in press).

Wagner, A. (2002). Asymmetric functional divergence of duplicate genes in yeast, *Molecular Biology and Evolution*, vol. 19, pp. 1760–1768.

Wagner, A. (2003). How the global structure of protein interaction networks evolves, *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, pp. 457–466.

Watts, DJ. & Strogatz, SH. (1998). Collective dynamics of 'small-world' networks, *Nature*, vol. 393, pp. 440–442.

Xu, J. & Li, Y. (2006). Discovering disease-genes by topological features in human protein-protein interaction network., *Bioinformatics*, vol. 22, pp. 2800–2805.

Yamanishi, Y.; Kotera, M.; Kanehisa, M. & Goto, S. (2010). Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework, *Bioinformatics*, vol. 26, pp. i246–254.

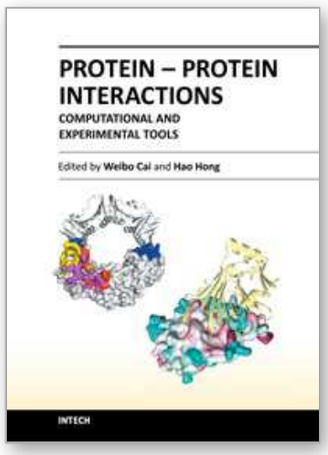
Yao, L. & Rzhetsky, A. (2008). Quantitative systems-level determinants of human genes targeted by successful drugs, *Genome Research*, vol. 18, pp. 216–213.

Yildirim, MA.; Goh, KI.; Cusick, ME.; Barabasi, AL. & Vidal, M. (2007). Drug-target network, *Nature Biotechnology*, vol. 25, pp. 1119–1126.

Yu, H.; Braun, P.; Yildirim, MA.; Lemmens, I.; Venkatesan, K.; Sahalie, J.; Hirozane-Kishikawa, T.; Gebreab, F.; Li, N.; Simonis, N.; Hao, T.; Rual, JF.; Dricot, A.; Vazquez, A.; Murray, RR.; Simon, C.; Tardivo, L.; Tam, S.; Svrzikapa, N.; Fan, C.; de Smet, AS.; Motyl, A.; Hudson, ME.; Park, J.; Xin, X.; Cusick, ME.; Moore, T.; Boone, C.; Snyder, M.; Roth, FP.; Barabási, AL.; Tavernier, J.; Hill, DE. & Vidal, M. (2008). High-quality binary protein interaction map of the yeast interactome network, *Science*, vol. 322, pp. 104–110.

Zanzoni, A.; Soler-Lopez, M. & Aloy, P. (2009). A network medicine approach to human disease, *FEBS Letters*, vol. 583, pp. 1759–1765.

Zhao, S. & Li, S. (2010). Network-based relating pharmacological and genomic spaces for drug target identification, *PLoS ONE*, vol. 5, e11764.



Protein-Protein Interactions - Computational and Experimental Tools

Edited by Dr. Weibo Cai

ISBN 978-953-51-0397-4
Hard cover, 472 pages
Publisher InTech
Published online 30, March, 2012
Published in print edition March, 2012

Proteins are indispensable players in virtually all biological events. The functions of proteins are coordinated through intricate regulatory networks of transient protein-protein interactions (PPIs). To predict and/or study PPIs, a wide variety of techniques have been developed over the last several decades. Many in vitro and in vivo assays have been implemented to explore the mechanism of these ubiquitous interactions. However, despite significant advances in these experimental approaches, many limitations exist such as false-positives/false-negatives, difficulty in obtaining crystal structures of proteins, challenges in the detection of transient PPI, among others. To overcome these limitations, many computational approaches have been developed which are becoming increasingly widely used to facilitate the investigation of PPIs. This book has gathered an ensemble of experts in the field, in 22 chapters, which have been broadly categorized into Computational Approaches, Experimental Approaches, and Others.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Takeshi Hase and Yoshihito Niimura (2012). Protein-Protein Interaction Networks: Structures, Evolution, and Application to Drug Design, Protein-Protein Interactions - Computational and Experimental Tools, Dr. Weibo Cai (Ed.), ISBN: 978-953-51-0397-4, InTech, Available from: http://www.intechopen.com/books/protein-protein-interactions-computational-and-experimental-tools/protein-protein-interaction-networks-structures-evolution-and-application-to-drug-design

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen