

PUBLISHED BY

# INTECH

open science | open minds

World's largest Science,  
Technology & Medicine  
Open Access book publisher



**2,850+**  
OPEN ACCESS BOOKS



**98,000+**  
INTERNATIONAL  
AUTHORS AND EDITORS



**91+ MILLION**  
DOWNLOADS



**BOOKS**  
DELIVERED TO  
151 COUNTRIES

AUTHORS AMONG  
**TOP 1%**  
MOST CITED SCIENTIST



**12.2%**  
AUTHORS AND EDITORS  
FROM TOP 500 UNIVERSITIES



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

Chapter from the book *Medicinal Chemistry and Drug Design*

Downloaded from: <http://www.intechopen.com/books/medicinal-chemistry-and-drug-design>

Interested in publishing with InTechOpen?  
Contact us at [book.department@intechopen.com](mailto:book.department@intechopen.com)

# Analysis of Protein Interaction Networks to Prioritize Drug Targets of Neglected-Diseases Pathogens

Aldo Segura-Cabrera<sup>1,5</sup>, Carlos A. García-Pérez<sup>1</sup>,  
Mario A. Rodríguez-Pérez<sup>2</sup>, Xianwu Guo<sup>2</sup>,  
Gildardo Rivera<sup>3</sup> and Virgilio Bocanegra-García<sup>4</sup>

<sup>1</sup>Laboratorio de Bioinformática

<sup>2</sup>Laboratorio de Biomedicina Molecular

<sup>3</sup>Laboratorio de Biotecnología Ambiental

<sup>4</sup>Laboratorio de Medicina de Conservación

Centro de Biotecnología Genómica, Instituto Politécnico Nacional

<sup>5</sup>U.A.M. Reynosa Aztlán, Universidad Autónoma de Tamaulipas, Reynosa  
México

## 1. Introduction

Many technological, social and biological systems have been modeled in terms of large networks providing invaluable insight in the understanding of such systems. Systems biology is an emerging and multi-disciplinary discipline that studies the interactions of cellular components by treating them as part of an integrated system. Thus, systems biology has shown that functional molecules are involved in complex networks of inter-relationships, and that most of the cellular processes depend on functional modules rather than isolated components. Large amounts of biological network data of different types are available, e.g., protein-protein interaction, transcriptional regulatory, signal transduction, and metabolic networks. Since proteins carry out most biological processes, the protein interaction networks (PINs) are of particular importance. The advancement of the functional genomics and systems biology of model organisms such as *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster* has contributed to the development of experimental and computational methods, and also to the understanding of human complex diseases. The availability of these methods has facilitated systematic efforts at creating large-scale data sets of protein interactions, which are modeled as PINs.

Usually, a PIN is represented as a graph where the proteins are the nodes and the interactions are the edges. According to the complex network theory, PINs are scale-free networks characterized by a power-law degree distribution. In scale-free networks, most nodes have a small number of links between them; whereas, a small percentage of nodes interact with a disproportionately large number of others. The nodes with a large number of links in PINs are called hub proteins. Functional genomics studies showed that in PINs, the deletion of a hub protein is lethal to the organism, a phenomenon known as the centrality-

lethality rule. This rule is widely believed to reflect the special importance of hubs in organizing the network, which in turn suggests the biological significance of network topology. Several well-known studied proteins that are implicated in human diseases are hub proteins. Examples include p53, p21, p27, BRCA1, ubiquitin, calmodulin, and others which play central roles in various cellular mechanisms.

Despite recent advances in systems biology of model organisms, the systems biology of human pathogenic organisms such as those that cause the so-called "neglected-diseases" has not received much attention. Neglected-diseases are chronic or related disabling infections affecting more than 1 billion people worldwide, mainly in Africa. Pathogens of neglected-diseases include: Protozoan parasites (e.g., *Leishmania* spp., *Plasmodium* spp., and *Trypanosoma* spp.), vector-borne helminthes (e.g., *Schistosoma* spp., *Brugia malayi*, and *Onchocerca volvulus*), soil-transmitted helminthes (e.g., *Ascaris lumbricoides* and *Trichuris trichura*), bacteria (e.g., *Mycobacterium tuberculosis* and *M. leprae*), and viruses (e.g., dengue and yellow fever virus). A number of factors limit the utility of existing drugs in neglected-diseases such as high cost, poor compliance, drug resistance, low efficacy, and poor safety. Since the evolution of drug resistance is likely to compromise every drug over time, the demand for new drugs and targets is continuous. The drug target identification is the first step in the drug discovery flow-through process. This step is complicated because a drug target must satisfy a variety of criteria. The important factors in this context are mainly related to the toxicity to host, and the essentiality of the target to the pathogen's physiology for growth and survival. Thus, the topological and functional analysis of neglected-disease pathogen PINs offers a potentially effective strategy for identifying and prioritizing new drug targets.

This chapter will introduce the reader to the basic concepts of network analyses and outline why it is important in terms of predicting protein function and essentiality. Work involving PINs of neglected-disease pathogens will be explained so that the reader will understand the current state in terms of its application to prioritize drug targets. The experimental and computational methods most likely to be used to identify and predict PINs, and the strategies for identifying multiple potential drug targets in neglected-disease pathogens will be also outlined using several biological databases in an integrated way.

To achieve this goal, the chapter includes three sections. Firstly, we present an outline of the conceptual development of network biology. The applied functional genomics involving the analysis of PINs of model organisms has led to developing methods and principles for elucidating protein function. We will also explain how these concepts are connected with protein essentiality to identify their "weak" points on the PINs of neglected-disease pathogens and its use for prioritizing drug targets. In the second section, we outline the experimental and computational methods that are most extensively to be used to identify and predict PINs. Some new approaches for predicting PINs are also introduced. These include the probabilistic integrated network methods which have shown the capability to increase the accuracy and coverage of the PINs. These primary research articles will be reviewed and the potential applications for the future be explained. This section mainly focused on analyzing the PINs of most prevalent neglected-disease pathogens in which the use of drugs is often limited by factors including high cost, low efficacy, toxicity, and the emergence of drug resistance. The potential use as an integrated strategy aimed at prioritizing and identifying drug targets of neglected-disease pathogens will be put forward, and the argument for future research involving the application of many tools and strategies will be discussed. In the final section,

we describe, amenably, the basic criteria to select pathogen drug targets, and the PINs of neglected-disease pathogens will be described in such a manner that the chapter will work as a source of key literature references for students and researchers. Papers will be reviewed to describe these basic principles, using key publications containing data and quantitative analyses (models, figures, tables) for PINs of some neglected-disease pathogens. We will describe novel lines of research; pros and cons of the use of PINs for prioritizing and identifying drug targets of neglected-disease pathogens.

## 2. Systems and network biology: Basic concepts

Systems biology is a holistic approach that involves the study of the inter-relationships of all the different elements in a biological system in order to understand non-deterministic behaviors that emerge from interaction between the cellular components and their environment and not by studying them in an isolated manner, one at a time (Hood and Perlmutter 2004, Weston and Hood 2004, Kohl and Noble 2009). Thus, the cell's behavior can be understood as a consequence of the complex interactions between its numerous constituents such as DNA, RNA, proteins, and metabolites. These interactions are also responsible for performing processes critical to cellular survival. For example, during transcription process the regulatory proteins can activate or inhibit the expression of genes or regulate each other as part of gene regulatory networks. Likewise, the cellular metabolism can be integrated into a metabolic network whose fluxes are regulated by enzymes. Similarly, the PINs represent how the proteins work together through interactions that lead to the modification of protein functions or new roles in protein complexes.

The biological systems consisting of interacting cellular components have led to the use of graph theory and mathematical tools based on graphs where the individual components are represented by nodes and the interactions by links (Fig. 1). Albert and Barabási (2002) have shown the general properties found among several networks ranging from the Internet to social and biological networks (Albert and Barabási 2002). The analysis of topology of those networks showed that they deviate substantially from randomly built networks as studied by Erdős and Rényi (Fig. 1a) (Erdős and Rényi 1960). Also, these networks did not show a well-shaped frequency distribution of the number of links per node as expected from randomly formed networks; instead, they showed a power-law distribution, which is characteristic of scale-free networks (Fig. 1b and 1c) (Amaral *et al.*, 2000, Albert 2005).

In scale-free network, the majority of nodes have only a few links, whereas very few nodes have a large number of links. Those nodes are called hubs and they represent the most vulnerable points of a network (Barabasi and Albert 1999, Albert *et al.*, 2000, Jeong *et al.*, 2001, Yu *et al.*, 2004a, Tew *et al.*, 2007). The topological features of networks can be quantified by measuring topological parameters whose information content provides a description from local (e.g., single nodes or links) to network-wide level (e.g., connections and relationships between nodes). For example, the nodes of a graph can be characterized by means of the number of links they have (the number of other nodes to which they are connected). This parameter is called "node degree". In directed networks, it is possible to distinguish the number of directed links that points toward the node (in-degree), and the number of directed edges that points outward the node (out-degree). The node degree characterizes individual nodes; however, in order to relate this parameter to whole network, a network degree distribution can be defined. The degree distribution  $P(k)$  represents the

fraction of nodes that have degree  $k$  and it is obtained by counting the number of nodes  $N(k)$  that have  $k = 1, 2, \dots$  links and dividing it by the total number of nodes  $N$ . The degree distributions of numerous networks such as the Internet, social, and biological networks, follow a power law (Fig. 1b and 1c) which is defined by the functional equation  $P(k) \sim k^{-\gamma}$ , where  $\gamma$  represents the degree exponent, taking usually values in the range between  $2 < \gamma < 3$  (Barabasi and Oltvai 2004). This function is intimately linked to the growth of the network in which new nodes are preferentially attached to already established nodes, a property that is also thought to characterize the evolution of biological systems (Jeong *et al.*, 2000).

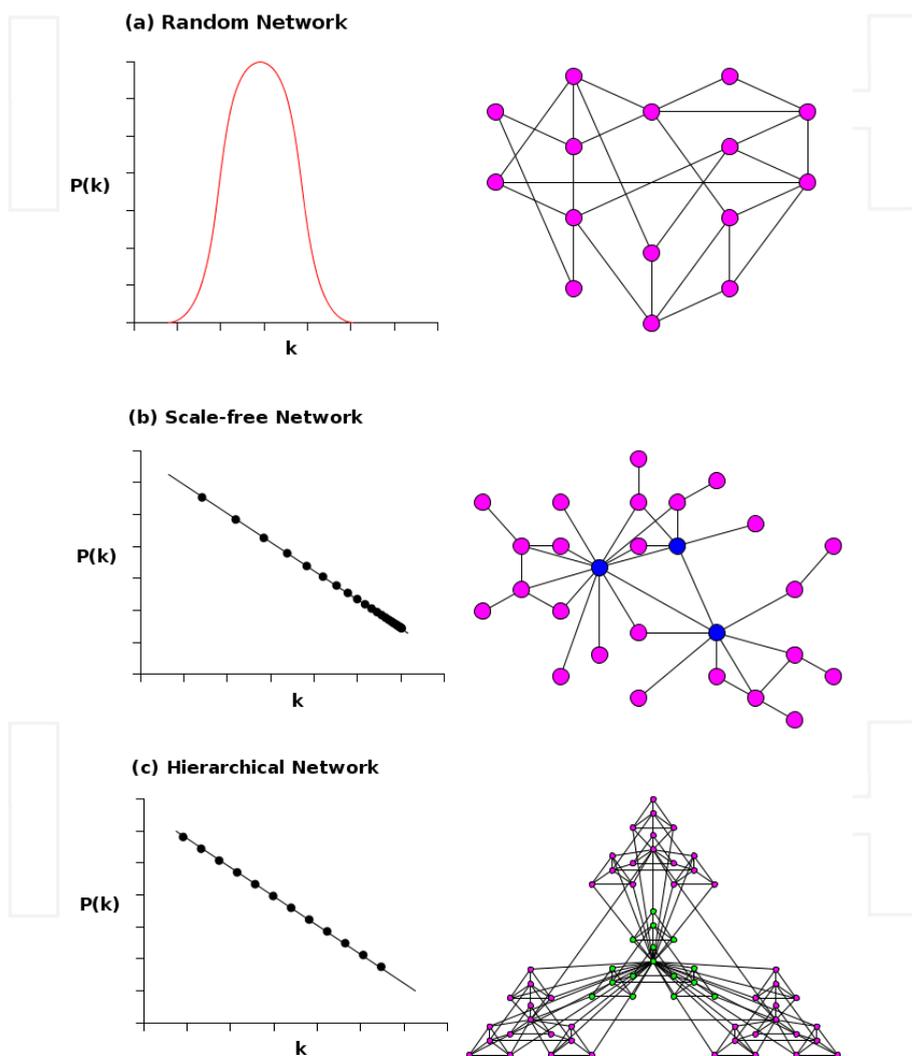


Fig. 1. Three types of network models and their associated distributions: (a) random network, (b) scale-free network, and (c) hierarchical network.

The distance between any two nodes in a network could be defined by the path length. In other words, it represents how many links we need to pass between two nodes. Nevertheless, it could have many alternative paths between two nodes in a network. The path with the smallest number of links between the selected nodes (shortest path) is of special interest. A common characteristic of several biological networks, including metabolic networks (Jeong *et al.*, 2000, Wagner and Fell 2001) and PINs (Giot *et al.*, 2003, Yook *et al.*, 2004) is that any two nodes can be connected with a path of a few links only. The main biological implications of this characteristic are related to: i) how the biological networks are capable of rapid responses to perturbations; ii) its capacity to employ alternative roads for the same input and output; and iii) the ability to efficiently compensate the perturbations in essential pathways.

Another important issue derived from network analysis is the concept of modularity, which can be used to describe how a group of physically or functionally linked nodes work together to achieve a particular function. The topological parameter used to quantify the modularity in a network is the clustering coefficient  $C_i$ , which represents the ratio between the number of links connecting nodes adjacent to node  $i$  and the total possible number of links among them (Watts and Strogatz 1998). It is worth noting that in first instance, the modularity concept might be in contradiction of the scale-free nature of the networks because the presence of modules implies that there are clusters of nodes that are relatively isolated from the rest of the network. However, it has been demonstrated that modularity and scale-free properties naturally co-occur in biological networks indicating that modules are not independent, instead, they are combined to form a hierarchical network (Fig. 1c) (Ravasz *et al.*, 2002).

Biological networks, including PINs and metabolic networks are good examples of network modularity because they exhibit high average  $C_i$ , which are associated to a high level of network robustness (Alon *et al.*, 1999, Ravasz *et al.*, 2002, Barabasi and Oltvai 2004). The most common representation of a module or cluster in a network is as a highly interconnected group of nodes. The biological implication of the modularity concept is that the nodes that integrate a module tend to participate in related biological processes and pathways; for example, protein and nucleic-acid synthesis, protein degradation, signal transduction, and metabolic pathways (Ma'ayan *et al.*, 2005). The analysis of experimental PINs have shown to have a remarkably modularity character (Giot *et al.*, 2003, Yook *et al.*, 2004). These findings in experimental PIN maps have been used to improve the understanding of the pleiotropic effects, and how perturbations on genes or proteins can propagate through the network and produce, in appearance, unrelated or extensive effects.

In addition to the modules, within a network, small and recurring sub-graphs, known as interaction motifs, with well-defined topologies can be identified (Fig. 2). The frequency analysis of these interaction motifs in networks revealed that they are over-represented when compared to a randomized version of the same network, suggesting that not all sub-graphs are equally significant in networks and that interaction motifs form functionally separable building blocks of cellular networks (Mangan and Alon 2003, Wuchty *et al.*, 2003, Alon 2007). For example, triangle motifs, also called feed-forward loops in directed networks, appear in both transcription-regulatory and neural networks. Likewise, there is evidence suggesting that specific motif type aggregates to form large motif clusters and that also appear to be commonly involved with certain functional roles (Milo *et al.*, 2002, Shen-

Orr *et al.*, 2002, Wuchty *et al.*, 2003). For example, in the *E. coli* transcription regulatory network, most motifs overlap, in which the specific motifs are no longer clearly separable (Shen-Orr *et al.*, 2002).

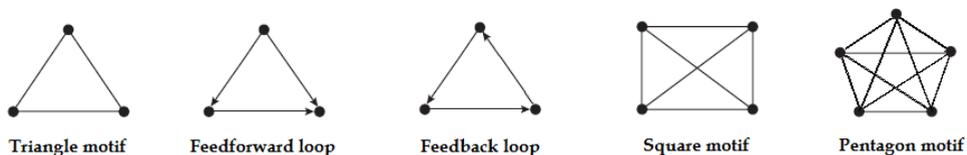


Fig. 2. Some types of interaction motifs found in biological networks.

The relevance of any node in mediating the communications flow among other nodes in the network is quantified by its betweenness centrality, which is defined as the total number of non-redundant shortest paths going through a certain node or edge (Freeman 1977). Girvan and Newman (2002), have proposed that the edges with high betweenness are the ones that are “between” network clusters; therefore, the information flow within a network could be altered by removing these edges (Girvan and Newman 2002). Dunn *et al.*, (2005) using an edge betweenness based-method have shown that clusters in PINs tend to share similar functions (Dunn *et al.*, 2005). Moreover, Yu *et al.*, (2007) have reconsidered the classical meaning of betweenness as a measure of the centrality of the nodes in a PIN. They have defined those nodes as “bottlenecks” with the highest betweenness centrality and find that bottlenecks nodes have a higher probability to be essential (Yu *et al.*, 2007).

It is worth noting that the topological parameters might be combined between them or with additional information of functional annotations regarding the network nodes (genes or proteins). Thus, a network provides testable predictions ranging from single interactions to essential genes and functional modules (del Rio *et al.*, 2009). Likewise, the functions of un-annotated genes or proteins can be also predicted on the basis of the annotation of their interacting partners. This approach to predict the protein/gene function is known as “guilty by association”. Additionally, the integration of information related to diseases or specific phenotypes with network approaches also enhances the understanding of human diseases, pharmacology response, and phenotype prediction (Ideker and Sharan 2008, Lee *et al.*, 2008a, Lee *et al.*, 2010, Wang and Marcotte 2010, Lee *et al.*, 2011).

### 3. Methods to identify protein interactions networks (PINs)

#### 3.1 Experimental methods

In the postgenomic era, the accumulation of protein-protein interaction data has enabled the biology systems studies at PINs levels (von Mering *et al.*, 2002). However, PIN analysis requires methods amenable to high throughput (HT) screening, such as large-scale versions of techniques like yeast two hybrid (Y2H) and tandem affinity purification coupled to mass spectrometry (TAP-MS) for performing systematic screens (Ito *et al.*, 2001a, Cusick *et al.*, 2005). In addition, there are a wide variety of methods to detect, analyze, and quantify protein interactions, including surface plasmon resonance spectroscopy, nuclear magnetic resonance (NMR), x-ray crystallography, and fluorescence-based technologies. These techniques provide detailed information on physical properties of protein interactions.

These methods are of paramount usefulness; however, herein, the techniques that can be applied to determine protein-protein interactions, at large-scale level, will be highlighted. In particular, the outcomes of Y2H system and TAP-MS are used further to perform *in silico* global network analysis. Both techniques were intensively applied to map the PIN of yeast, the first model organism with available PINs (Uetz *et al.*, 2000, Ito *et al.*, 2001b, Gavin *et al.*, 2002, Ho *et al.*, 2002, Ito *et al.*, 2002, Tong *et al.*, 2004, Yu *et al.*, 2008). Afterwards, large-scale efforts have been made to determine PINs for other model minor eukaryotic organisms: *D. melanogaster* (Giot *et al.*, 2003), and *C. elegans* (Li *et al.*, 2004); pathogenic microorganisms: *Helicobacter pylori*, *Campylobacter jejuni*, *Treponema pallidum*, *M. tuberculosis* (Wang *et al.*, 2010), herpes simplex virus 1 (Lee *et al.*, 2008b), and Kaposi's sarcoma-associated herpesvirus (Uetz *et al.*, 2006, Rozen *et al.*, 2008), and major eukaryotic organisms: *Arabidopsis thaliana* (de Folter *et al.*, 2005) and humans (Rual *et al.*, 2005, Stelzl *et al.*, 2005, Gandhi *et al.*, 2006). Even though the PINs are not completed, the available PINs provide insight into how particular properties of proteins are integrated at systems level, and also, as a useful resource to predict the functional role of genes or proteins.

### 3.1.2 Yeast two-hybrid (Y2H) system

The Y2H system has considerably accelerated the *in vivo* large-scale screening of protein interactions enabling the detection of physically interacting proteins by using the modular organization of eukaryotic transcriptional activators. The eukaryotic transcription activators are formed by at least two distinct domains, one responsible of binding to a DNA region (BD) promoter and the other of activating the transcriptional processes (AD). It is well-known that splitting BD and AD domains will inactivate the transcriptional processes, but the transcription can be restored if a BD domain is re-associated with an AD domain (Fields and Song 1989). Thus, the standard Y2H system includes a DB domain fused to the "bait" protein-coding region and an AD domain fused to the "prey" protein-coding region. When DB-bait and AD-prey domains are co-expressed in the nucleus of yeast cells, "bait"- "prey" domain interaction reconstitutes a functional transcription factor that activates the transcription of one reporter gene (Fig. 3). The most used Y2H system is based on GAL4/LexA, where the GAL4 protein controls the expression of the LacZ gene encoding beta-galactosidase.

The main advantages of Y2H system are: i) the DNA ( not the protein) is manipulated to study both bait and prey proteins (Walhout and Vidal 2001a); ii) it allows to identify protein interactions *in vivo*; iii) to identify transitory protein interactions, and iv) it is amenable to high-throughput screening methods (Buckholz *et al.*, 1999, Uetz and Hughes 2000, Walhout and Vidal 2001b, Ito *et al.*, 2002, Rual *et al.*, 2005).

The drawbacks include: i) a high proportion of false-positives and negatives (Vidal and Legrain 1999, Ito *et al.*, 2002); ii) it forces sub-cellular localization of bait and prey in the yeast nucleus which might preclude certain interactions from taking place (Cusick *et al.*, 2005). For example, membrane protein interactions cannot be identified by standard Y2H system because the AD-prey fusion will be retained at the membrane, thus, avoiding the reconstitution of a functional transcription factor (Xia *et al.*, 2006); iii) the over-expression of tested proteins, thus modifying the relative concentrations of potential interaction partners in comparison to the *in vivo* state; iv) the presence of auto-activators, i.e. proteins initiating

transcription by themselves (Cusick *et al.*, 2005), and v) the differences in post-translational modifications and protein folding processes between yeasts and other organisms (Shoemaker and Panchenko 2007). Given these cons, several modifications have been made to improve the quality of the Y2H system results, including the development of membrane Y2H, the inclusion of different promoters of reporter genes, the use of low copy vectors, and the reduction of auto-activators. Once that these drawbacks are reduced, the quality of the Y2H system is significantly improved (Lehner *et al.*, 2004, Li *et al.*, 2004, Rual *et al.*, 2005, Yu *et al.*, 2008).

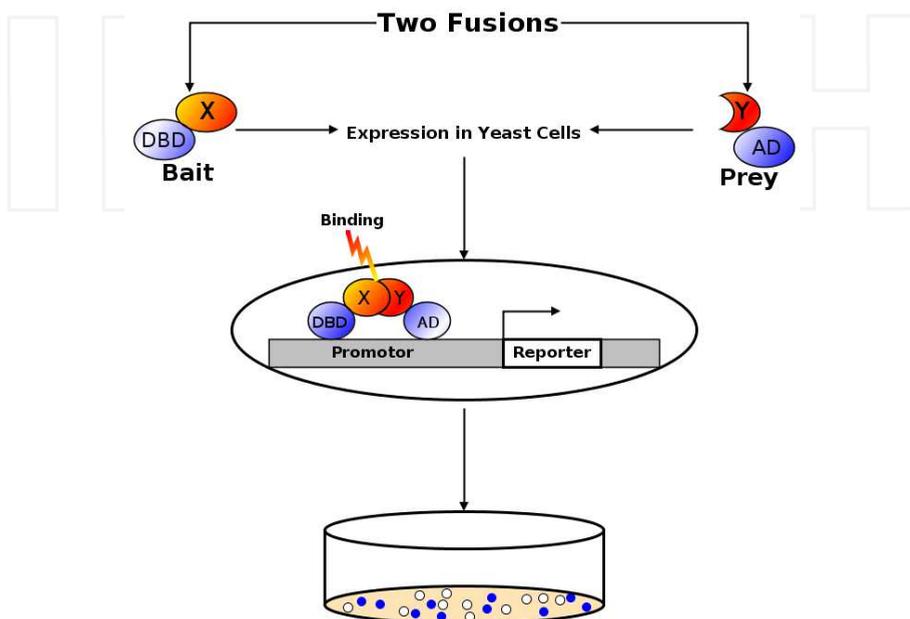


Fig. 3. The Y2H system. Y2H detects interactions between proteins X and Y, where X is linked to BD domain which binds to DNA region promoter.

### 3.1.3 Tandem affinity purification-tag coupled to mass spectrometry (TAP-MS)

TAP-MS method is a powerful approach to determine the composition of relevant protein complexes. In this method, a target protein-coding region is fused with a DNA sequence encoding an affinity tag which will be expressed with other cellular proteins, followed by two-step affinity purification (AP) and elucidation of the complex components by mass spectrometry (MS). A typical TAP tag is formed by an immunoglobulin interacting domain of protein A (protA) and a calmodulin-binding peptide (CBP) (Fig. 4). The protA/CBP binding domains are separated by a short recognition sequence for the site-specific tobacco etch virus protease (TEV protease). The TEV site allows proteolytic elution of the protein complex from IgG-sepharose after the first affinity-purification step, which is based on the protA/IgG-sepharose interaction. The eluted protein complex is further purified by binding to a calmodulin affinity resin, eluted with EGTA and processed for identification with MS analyses.

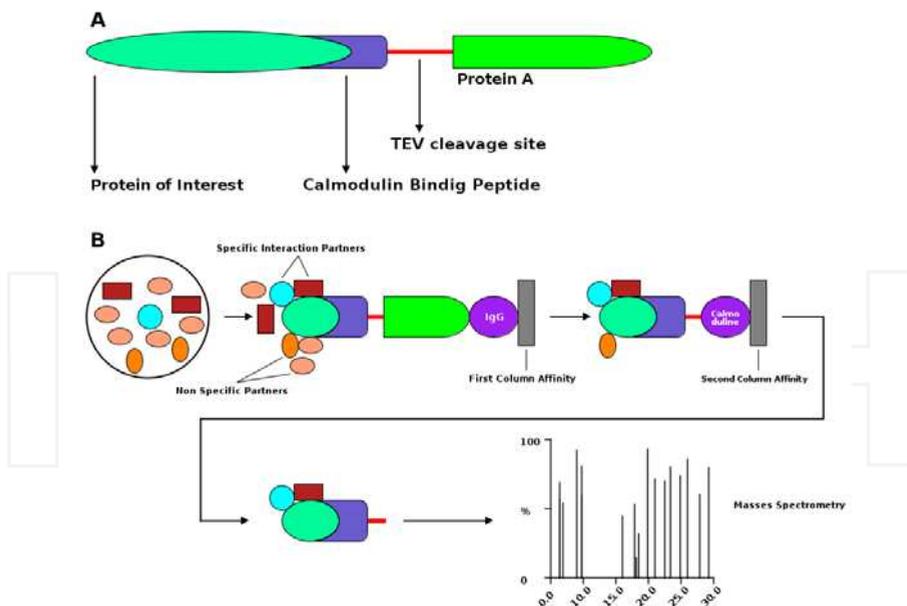


Fig. 4. TAP-MS method. TAP purifies protein complexes and removes the molecules of contaminants and MS identifies the complex components.

Similar to Y2H system results, TAP-MS method shows a high rate of false-positives and negatives, missing many transient interactions. In contrast to the Y2H system, the TAP-MS method can elucidate higher-order interactions beyond binary interactions and, therefore, provides direct information on protein complexes. Several large-scale studies of protein complexes have been performed using the TAP-MS method (Gavin *et al.*, 2002, Ho *et al.*, 2002, Gavin *et al.*, 2006). For example, Gavin *et al.*, (2006) used 5,500 ORFs fused to DNA sequences encoding an affinity tag to analyze PIN of *S. cerevisiae*. They found 491 complexes, of which 257 are novel, showing that PIN in *S. cerevisiae* has a modular organization (Gavin *et al.*, 2006). In addition, Stingl *et al.*, (2008), have elucidated the urease interactome of *H. pylori*. They combined the tandem affinity purification protocol with *in vivo* cross-link in order to capture transient interactions, which represent an improvement to TAP-MS method (Stingl *et al.*, 2008).

The use of experimental orthogonal approaches has demonstrated that Y2H and TAP-MS interaction data sets contain mostly highly reliable interactions. It has been suggested that the integration of data from the two approaches can also serve to increase confidence in either data set, and has provided support to derivate predictions from these approaches (Cusick *et al.*, 2005). Moreover, Venkatesan *et al.*, (2009) have developed a framework to estimate various quality parameters associated with currently used methods to identify PINs. The combination of these quality parameters (screening completeness, assay sensitivity, sampling sensitivity, and precision), has shown an estimate of the size of human binary interactome and a path toward the completion of its mapping (Venkatesan *et al.*, 2009).

Despite the technical or biological limitations (Cusick et al., 2005) of the aforementioned methods, that does not preclude a reduction on their impact in PINs studies, instead they are marking a paradigm change from one-gene/one-function reductionist approach to a more systemic approach that can capture all potential interactions encoded in a genome or proteome.

### 3.1.4 Protein interaction databases

The huge amounts of protein interaction data produced by high-throughput experimental methods as Y2H and TAP-MS and analyzed by bioinformatics have led to the conformation of several research groups aimed at conducting important efforts in designing and setting up databases that include carefully analyzed information to provide useful scientific knowledge about protein-protein interactions. Table 1 shows a summary of most significant public databases of protein-protein interactions published to date. These databases contain interactions obtained by direct submission from experimentalists, text-mining and other data sources. Also, there are other online resources integrating information from several of the databases that are listed in Table 1, or tools to browse and visualize such data; for example resources like APID (Prieto and De Las Rivas 2006, Hernandez-Toro *et al.*, 2007) and PINA (Wu *et al.*, 2009). The information deposited in these databases is verified using automated algorithms or manual curation like in the DIP database (Deane *et al.*, 2002). Altogether, protein interaction databases are an invaluable resource to develop projects that aims to analyze PINs of organisms ranging from viruses to humans.

Database	Type of data	Number of interactions	Website
DIP	E,C,S	71,589	<a href="http://dip.doe-mbi.ucla.edu">http://dip.doe-mbi.ucla.edu</a>
MINT	E,C	235,635	<a href="http://mint.bio.uniroma2.it">http://mint.bio.uniroma2.it</a>
IntAct	E,C	275,144	<a href="http://www.ebi.ac.uk/intact/">http://www.ebi.ac.uk/intact/</a>
BioGRID	E,C	282,005	<a href="http://thebiogrid.org/">http://thebiogrid.org/</a>
HPRD	E,C	39,194	<a href="http://www.hprd.org/">http://www.hprd.org/</a>
APID	I	322,579	<a href="http://bioinfow.dep.usal.es/apid/apid2net.html">http://bioinfow.dep.usal.es/apid/apid2net.html</a>
PINA	I	221,702	<a href="http://cbg.garvan.unsw.edu.au/pina">http://cbg.garvan.unsw.edu.au/pina</a>

Table 1. Most representative databases of protein-protein interactions. (E) high-throughput experimental data; (S) structural data; (C) manual curation, and (I) integrative resource. The number of interactions was updated on September 29, 2011.

### 3.2 Computational methods to predict protein interactions networks (PINs)

Parallel to the experimental methods, several computational methods have been designed to predict protein-protein interactions. Initially, these methods were strictly limited to proteins whose three-dimensional structures had been determined (structure-based methods). The completion of genome sequences has provided large amounts of genomic information enabling the analysis from a genomic context of a given gene. Thus, a number of

computational methods and resources have been developed for the prediction of protein interactions resulting from genomic information (genomic context-based methods), even in those cases where the three-dimensional structures are unknown yet (Galperin and Koonin 2000, Huynen *et al.*, 2000, Huynen and Snel 2000).

Hereinafter, we will describe computational methods and resources available for protein interaction prediction that exploit the genomic and biological contexts of proteins for complete genomes.

### 3.2.1 Genomic context-based methods

#### 3.2.1.1 Gene neighborhood

The gene neighborhood method exploits the notion that genes which physically interact or are functionally associated to the same process or functional pathway will be adjacent to each other in the genome (Fig. 5a) (Tamames *et al.*, 1997, Overbeek *et al.*, 1999, Bowers *et al.*, 2004). For example, Dandekar *et al.* (2005), have shown that the neighborhood relationship could be used as fingerprint, suggesting that the proteins encoded by these genes may physically interact (Dandekar *et al.*, 1998). The most representative example of this phenomenon can be found in bacterial operons, where genes that work together are generally transcribed as a unit. Furthermore, operons which encode for co-regulated genes

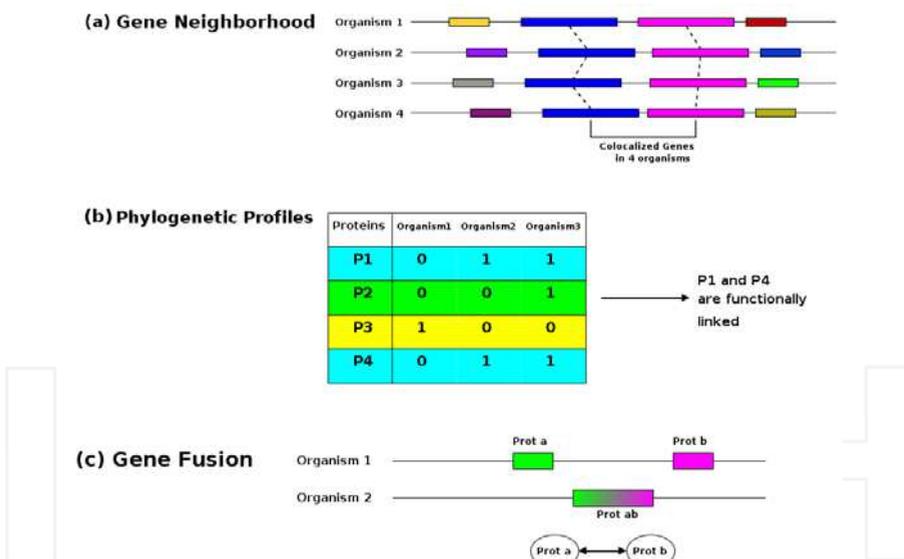


Fig. 5. Genomic context-based methods. (a) Gene neighborhood plots for four organisms, showing a pair of genes (blue and magenta) which are in close proximity in all four organisms. (b) Example phylogenetic profiles of four proteins from the three organisms. The proteins 1 and 4 have the same patterns of co-occurrence in all three organisms, and may physically interact based on this evidence. (c) A gene fusion event between two proteins (green and magenta) in two organisms is shown. Thus, the proteins a y b from organism 1 is predict to interact because they form part of a single protein in organism 2.

are usually conserved. The neighborhood relationship tends to be more relevant when it is conserved across different species (Tamames *et al.*, 1997). Hence, the gene neighborhood method, like many of the comparative genomics approaches, increases its robustness when a larger numbers of genomes are used for the prediction. Since operons and genes neighborhood are uncommon in eukaryotic species (Zorio *et al.*, 1994, Blumenthal 1998, Liu and Han 2009, Fitzpatrick *et al.*, 2010), this method is principally applicable to bacteria where such genome properties are relevant.

### 3.2.1.2 Phylogenetic profiles

The phylogenetic profile method is based on the co-occurrence of pairs of genes across multiple genomes (Fig. 5b). Consequently, a pair of orthologous genes remains together across many distant species representing a concerted evolution mechanism and indicating that these genes need to be simultaneously present to participate in the same biological process, pathway or physically interacting. A phylogenetic profile is commonly represented as a vector for the presence or absence of a gene across multiple genomes (Fig.), where "0" or "1" denoted the presence/absence at each position of a profile (Ouzounis and Kyrpides 1996, Rivera *et al.*, 1998, Pellegrini *et al.*, 1999).

The main drawbacks of this method are: it can only be applied to complete genomes; the prediction robustness is dependent on the number and distribution of genomes used to build the profile, thus, a pair of genes with similar profiles across many bacterial, archaeal and eukaryotic genomes is much more likely to interact each other than those genes found to co-occur in a small number of closely related species; its high computational cost since it needs to compare many complete genomes; and, fails in homology detection between distant organisms.

Like others genomic context methods, with the increasing number of completely sequenced genomes, it is expected that the accuracy of these predictions will be improved over time.

### 3.2.1.3 Gene fusion

The gene fusion method is based on the fact that some interacting protein domains (termed the rosetta stones) have homologs in other genomes that are fused into one protein chain (Fig. 5c). Thus, gene fusion events have been proposed for the identification of potential protein-protein interactions, metabolic or regulatory networks (Sali 1999, Galperin and Koonin 2000). The information about gene fusion events can be combined with phylogenomic profiling and identification of conserved chromosomal localization, to test hypotheses leading to the characterization of proteins of unknown function (Marcotte *et al.*, 1999a, Marcotte 2000, Enright and Ouzounis 2001). Marcotte *et al.*, (1999) found 6,809 potentially interacting pairs of non-homologous proteins in *E. coli*, revealing that, for more than half of the pairs, both involved members were functionally associated. More approaches with similar results have been used, including in eukaryotic genomes (Enright and Ouzounis 2001).

The drawbacks of this method are related with the domain complexity of eukaryotic proteins, the presence of promiscuous domains, and large degrees of paralogy (Enright *et al.*, 2002).

Currently, there are excellent resources implementing the genomic context-based methods. The most notable are the Search Tool for the Retrieval of Interacting Genes/Proteins

(STRING) and ProLinks. The STRING (URL: <http://string-db.org>) and ProLinks (URL: <http://prl.mbi.ucla.edu>) resources provide a web interface giving comprehensive access to gene context information in 1,100 and 900 complete genomes, respectively (Szkarczyk *et al.*, 2011, Bowers *et al.*, 2004).

### 3.2.2 Interologs

The use of homology relationships is a key paradigm in molecular biology and genomics. This approach has been extensively exploited to predict protein structure (Abagyan and Batalov 1997, Brenner *et al.*, 1998, Rost 1999), to study sub-cellular localization (Nair and Rost 2002), enzymatic activity (Devos and Valencia 2001, Todd *et al.*, 2001), and for comparative genomics (Marcotte *et al.*, 1999b, Pellegrini *et al.*, 1999). Thus, interologs is defined as a conserved interaction between a pair of proteins of a given organism which have interacting homologs in another organism (Yu *et al.*, 2004b). For example, the experimental observation that two yeast proteins interact is extrapolated to predict that the two corresponding homologs in human also interact in a similar way. Walhout (Walhout and Vidal 2001b) and Vidal (2001) have used yeast experimental interaction data (Uetz *et al.*, 2000, Ito *et al.*, 2001b) to infer similar interactions in worm (Fig. 6). Mika and Rost (2006) suggested that the extrapolation of interactions between distant organisms has to be undertaken with some caution. They found that the homology transfers are only accurate at high levels of sequence identity, and it is more reliable for protein pairs from the same species than for two protein pairs from different organisms (Mika and Rost 2006). Likewise, Wiles *et al.*, (2010) have developed a scoring schema to assess the confidence of interologs prediction. They have predicted protein interactions across five species (human, mouse, fly, worm, and yeast) based on available experimental evidence and conservation across species (Wiles *et al.*, 2010). Also, they developed the Interolog Finder (URL: <http://www.interologfinder.org>) to provide access to these data.

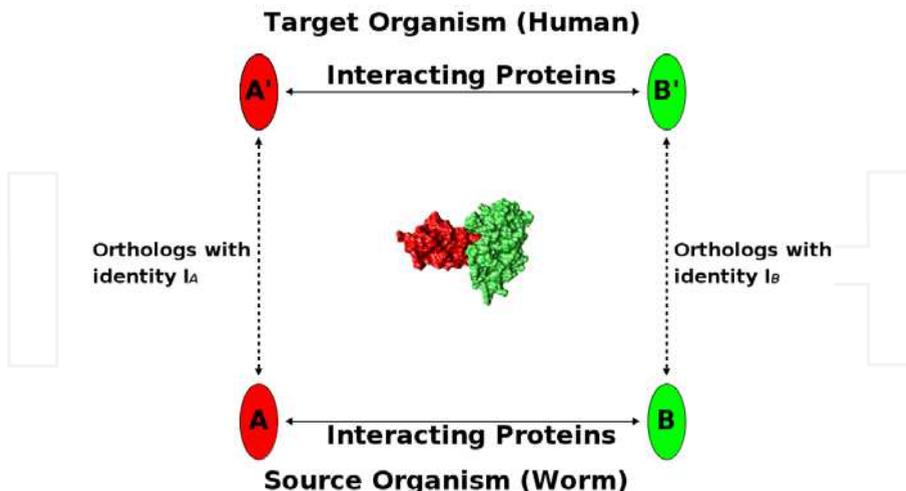


Fig. 6. The Interolog method. The A and B are interacting proteins in worm, and A' and B' are homologs in human of A and B proteins. Then A' and B' in human also interact in a similar way.

### 3.2.3 Integrative approaches

Currently, high-confidence PINs data sets are limited; however, they still provide a framework onto which other types of biological information can be integrated. Thus, new approaches that integrate other types of data, including protein-protein interactions, text mining, homology-based, and functional genomics approaches (Lee *et al.*, 2004, Chua *et al.*, 2007, Lee *et al.*, 2008a, Pena-Castillo *et al.*, 2008, Linghu *et al.*, 2009, Lee *et al.*, 2010, Wu *et al.*, 2010, Lee *et al.*, 2011, Szklarczyk *et al.*, 2011), have shown to be the most effective way to assign function to uncharacterized proteins that are components of the network (Fig. 7).

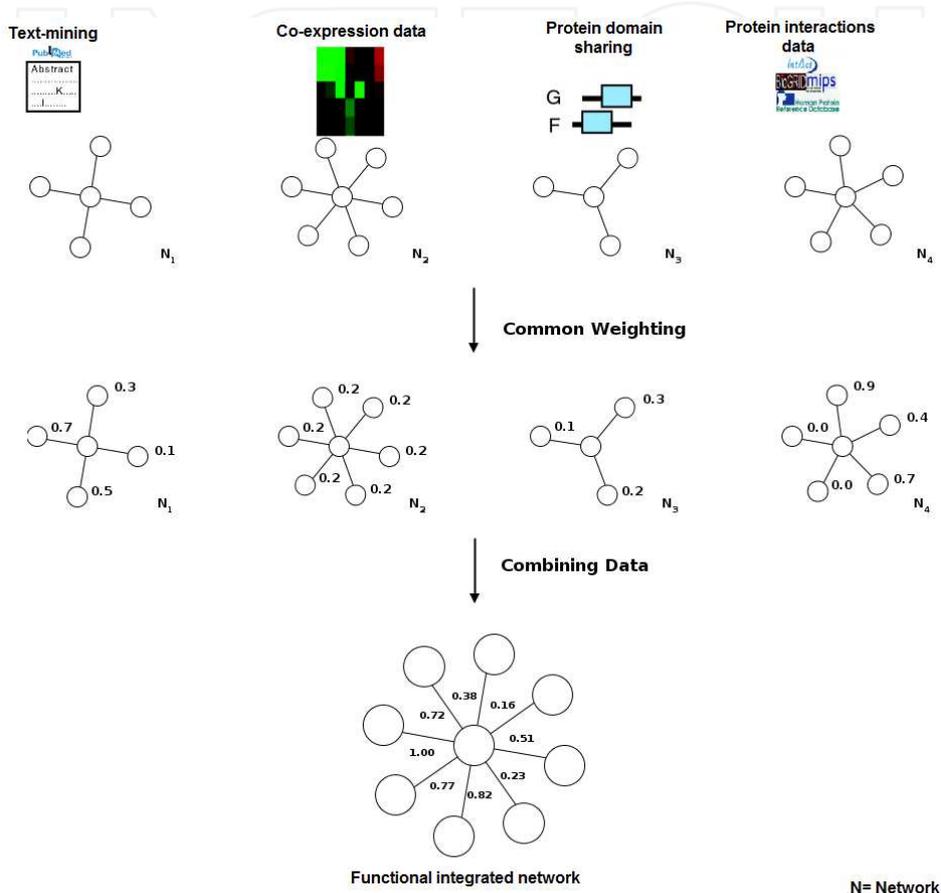


Fig. 7. General scheme for integrative approaches.  $N_1$ ,  $N_2$ ,  $N_3$  and  $N_4$  are networks representing four data sources. Each node is a protein, while each edge is a binary relationship. The edges are weighted into common weight that is consistent across different data sources.  $N_1$ ,  $N_2$ ,  $N_3$  and  $N_4$  are then combined and re-scored to form the final high confidence network  $N'$ .

The most representative example of these approaches is STRING which integrates experimental as well as predicted interaction information, mostly from the methods

aforementioned. STRING provides ease of access to explore this integrated information (URL: <http://string-db.org>). Moreover, for each protein-protein interaction it provides a confidence score, and supplementary information such as protein domains and 3D structures, all within a stable and consistent identifier space. The version 9.0 of STRING includes the information of more than 1,100 completely sequenced organisms, ranging from bacteria and archaea to humans allowing to periodically execute interaction prediction algorithms and update such data depending on genome sequence information (Szklarczyk *et al.*, 2011).

Similarly, several groups have integrated multiple networks to predict protein functions, interactions and functional modules including data from multiple sources, ranging from co-expression patterns, sequence similarity to genomic context-based methods (Kemmeren *et al.*, 2002, Jansen *et al.*, 2003, Lee *et al.*, 2004, Lu *et al.*, 2005, Chua *et al.*, 2007, Lee *et al.*, 2008a, Pena-Castillo *et al.*, 2008, Linghu *et al.*, 2009, Lee *et al.*, 2010, Wu *et al.*, 2010, Lee *et al.*, 2011). For example, Marcotte's group have shown the predictive power of an integrated functional network for *C. elegans* (Lee *et al.*, 2008a). Firstly, they computationally built an integrated functional network covering approximately 82% of *C. elegans* genes. Second, they used this network to predict the effects of perturbing individual genes on the organism's phenotype, identifying genes causing specific phenotypes ranging from cell cycle defects in single embryonic cells to life-span alterations, neuronal defects, and altered patterning of specific tissues. They select a set of candidate genes and their interactions associated to a phenotype and used RNAi to test whether targeting these candidate genes suppressed such phenotype. They found that 20% of such interactions suppressed the studied phenotype; instead, using only an RNAi, at large-scale screening, inactivation of 0.9% of genes produces such effect. Therefore, predictions arising from interactions of integrated network are 21-fold better than those expected by chance. They suggested a network-guided schema to accelerate research by using screening methods to identify genes and interactions for pathways of interest in human diseases.

The main limitation of integrative approaches is related with the availability of functional association data of genes/proteins. For example, these methods will not be able to make extensive predictions if no associations are available, as in the case of a novel genome with no known sequence or domain homology with known sequences, poorly studied genomes, and lack of functional genomics studies.

## **4. PINs as a tool to prioritize drug targets of neglected-disease pathogens**

### **4.1 Drug targets prioritization**

Despite the advent of the high-throughput techniques sparked by the genomics revolution, discovery and development of new drugs for neglected-disease pathogens has lagged in recent years due to the serious problems such as high cost, poor compliance, low efficacy, poor safety, evolution of antibiotic resistance, among others (Schmid 1998).

Target identification is the first step in the drug discovery process and such task can provide the foundation for years of dedicated research in the pharmaceutical industry (Read *et al.*, 2001). As compared with all the other steps in drug discovery, this stage is complicated by the fact that the identified drug target must satisfy a variety of criteria to permit progression to the next step. For example, the target must be selectively present in the pathogen, i.e.

target coding genes that are conserved across different pathogens and have no human homologs represent attractive target candidates for new broad-spectrum drugs (Schmid 2006); relevant for the pathogenesis process (Galperin and Koonin 1999, Sakharkar *et al.*, 2004); and, the essentiality of the target to the pathogen's growth and survival (Koonin *et al.*, 1998, Thanassi *et al.*, 2002, Galperin and Koonin 2004); suitability of the target for expression and assayability, and the availability of structures or models to initiate rational drug design (Aguero *et al.*, 2008). Hence, the integrated uses of above-mentioned strategies are considered as the basic schema in the drug target prioritization approaches. The criteria values of this basic schema can be found by querying publicly available bioinformatics resources and databases. For example, using metabolic pathway databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata *et al.*, 1999, Kanehisa and Goto 2000), protein classification sets such as Clusters of Orthologous Groups (COGs), Gene Ontology (GO), and resources to evaluate the "druggability" of proteins (Hopkins and Groom 2002, Russ and Lampel 2005, Hambly *et al.*, 2006), like "Structure-based DrugEBility" online service at EBI (URL: <https://www.ebi.ac.uk/chembl/drugability/structure>). For drug targets of neglected-disease pathogens, the TDR Targets Database (URL: <http://tdrtargets.org>) is an extensive resource for neglected tropical diseases (Aguero *et al.*, 2008). This database includes extensive genetic, biochemical, and pharmacological data related to tropical disease pathogens and computationally predicted druggability for potential targets. The database contains the data on the tuberculosis pathogen *M. tuberculosis*; the leprosy pathogen *M. leprae*; the malaria parasites *Plasmodium falciparum* and *P. vivax*, the toxoplasmosis parasite *Toxoplasma gondii*; the trematode *Schistosoma mansoni*; the filariasis helminth *Brugia malayi* and its intracellular symbiont bacterium *Wolbachia*; and the kinetoplastid parasites *Leishmania major*, *Trypanosoma brucei*, and *T. cruzi*, which are responsible for kala-azar and other forms of leishmaniasis, sleeping sickness, and Chagas disease, respectively.

## 4.2 PINs, drug targets, and neglected-disease pathogens

Networks analysis is a broadly applicable tool for the drug discovery and development process. Any type of association data linking one gene to another, a protein or a compound, can be modeled, visualized and analyzed as networks (Lee *et al.*, 2004, Chua *et al.*, 2007, Lee *et al.*, 2008a, Linghu *et al.*, 2009, Lee *et al.*, 2010, McGary *et al.*, 2010, Wu *et al.*, 2010, Lee *et al.*, 2011). Hence, data from pre-clinical and clinical trial studies can be included in network analyses (Nikolsky *et al.*, 2005). Thus, networks could represent the standard for data integration and analysis. Network analysis involving neglected-disease pathogens is a very young area of research. Moreover, despite the availability experimentally PINs of model organisms as *S. cerevisiae*, *C. elegans*, and *D. melanogaster*, and some bacterial pathogens like *H. pylori*, *C. jejuni*, *Treponema pallidum*, the number of experimentally neglected-disease pathogens PINs is limited. For example, LaCount *et al.*, (2005) identified protein-protein interactions of *P. falciparum* through a high throughput screening version of the yeast two-hybrid system (LaCount *et al.*, 2005). They found 2,846 unique interactions in more than 32,000 *P. falciparum* protein fragments. In order to determine clusters of interacting proteins they used computational methods such as analysis of network connectivity, gene co-expression, and enrichment of Gene Ontology terms. The results of the network analysis was the identification of two protein clusters, one of which related to the chromatin modification, transcription, messenger RNA stability, and ubiquitination and the other

implicated in the invasion of host cells. They suggested that the information provided by this network may be relevant to understand the basic biology of the parasite and to discover new drug and vaccine targets. Wang *et al.*, (2010) built a PIN of the *M. tuberculosis* H37Rv strain based on a high-throughput bacterial two-hybrid method. They found more than 8,000 novel interactions and performed a cross-species PINs comparison, showing 94 conserved sub-networks between *M. tuberculosis* and several prokaryotic PINs (Wang *et al.*, 2010).

Additionally, even the lack of data, several computational studies aims to predict PINs of neglected-disease pathogens and prioritize drug targets have been performed. Florez *et al.*, (2010) built an *in silico* PIN of *L. major* by combining information of PSIMAP, PEIMAP, iPfam databases, and using the interologs method (Florez *et al.*, 2010). They predicted 33,861 interactions for 1,366 proteins, and also analyzed the PIN by calculating topology parameters such as connectivity and betweenness centrality detecting 142 potential and specific drug targets without human orthologs (Fig. 8). Pedomallu and Posfai (2010) have developed a simple open source package module (OpenPPI\_predictor) to predict putative PIN for target genomes (Pedomallu and Posfai 2010). The package is based on interologs method and uses experimental data from a related organism. Thus, they assayed OpenPPI\_predictor to infer a PIN for *B. malayi* using experimental PIN data from *C. elegans*. They identified 118 and 143 clusters in *B. malayi* and *C. elegans* interactomes,

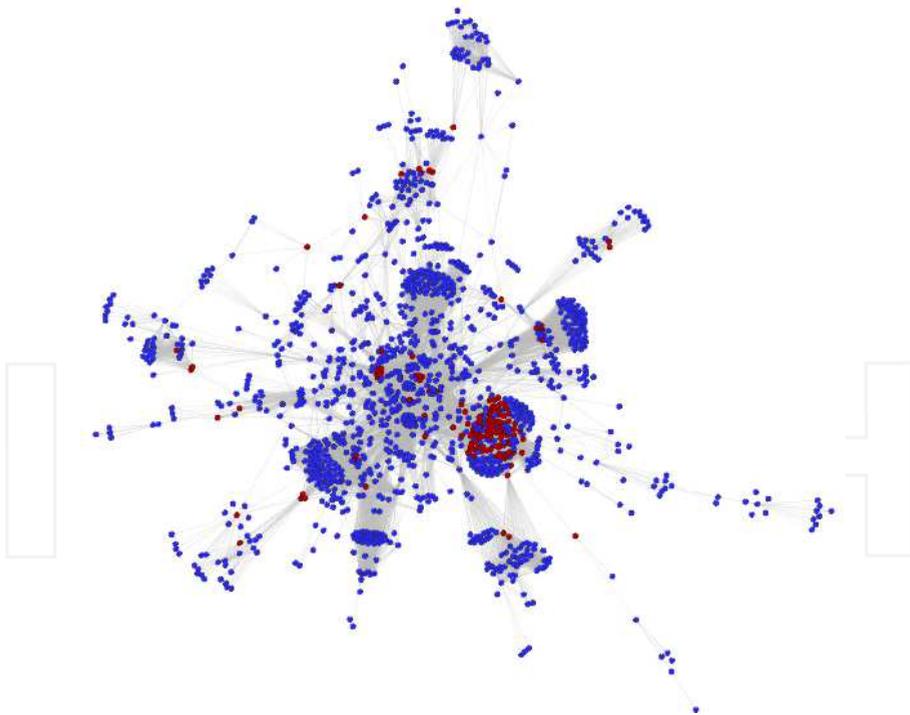


Fig. 8. Predicted PIN of *Leishmania major* by Florez *et al.*, (2010). The nodes in color red represent predicted essential proteins without human orthologs.

respectively, and found that highly connected region contains 363 and 340 proteins in *B. malayi* and *C. elegans* PINs. They suggests that core cellular functions of the two related organisms have similar complexity and that further analysis of these highly connected regions may provide clues about genes missing from a conserved pathway, or proteins missing from a complex.

Similarly, computational studies have been developed in order to model host-neglected-disease pathogens PINs. For example, Dyer *et al.*, (2007) integrated public intra-species PINs datasets with protein-domain profiles to predict a Human-*P. falciparum* PIN. They found 516 protein interactions between these two organisms, and showed that *Plasmodium* proteins interacting with human proteins are co-expressed in DNA microarray datasets, associated with developmental stages of the *Plasmodium* life cycle (Dyer *et al.*, 2007). Dyer *et al.*, (2008) have analyzed the landscape of human proteins interacting with pathogens. They integrated human-pathogen PINs for 190 pathogen strains from seven public databases and found that both viral and bacterial pathogens tend to interact with proteins with many interacting partners (hubs) and those that are central to many paths (bottlenecks) in the human PIN (Dyer *et al.*, 2008). Similar results were obtained by Navratil *et al.*, (2011). They used a high-quality dataset manually curated and validated of virus-host protein interactions to depict the “human infectome” (Navratil *et al.*, 2011). Additionally, they showed, by using functional genomic RNAi data, that the high centrality of targeted proteins was correlated to their essentiality for viruses’ lifecycle. Also, they perform a simulation of cellular network perturbations and showed a stealth-attack of viruses on proteins bridging cellular functions, which is a property that could be essential in the molecular etiology of some human diseases (Fig. 9). Doolittle and Gomez (2011) have predicted interactions between dengue

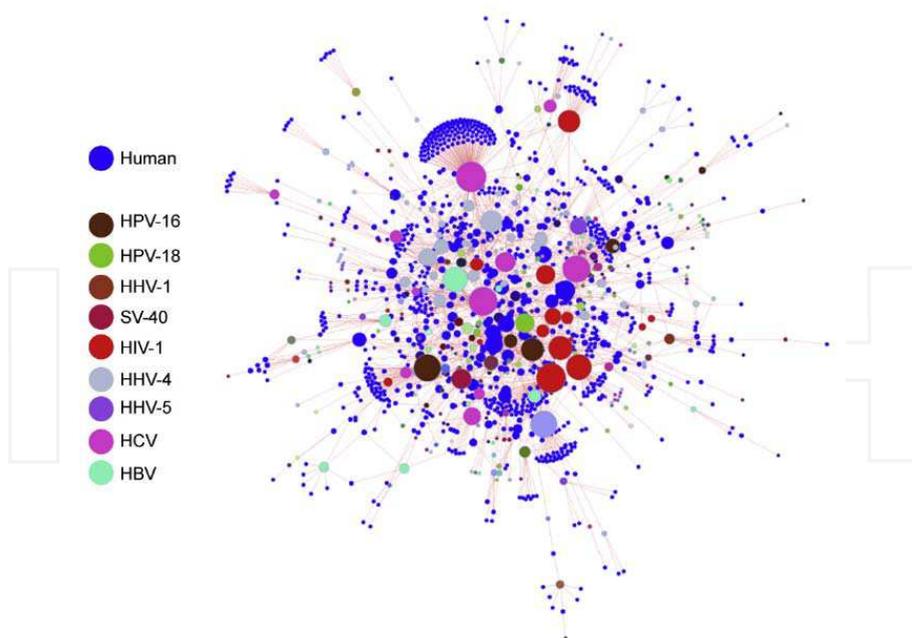


Fig. 9. The human infectome by Navratil *et al.*, (2011).

virus (DENV) and its hosts, both human and the insect vector *Aedes aegypti*. They implemented a protocol based on structural similarity between DENV and host proteins, and also they supported a subset of the predictions via mining from the literature. They predicted, after filtering and based on shared Gene Ontology cellular component, over 2,000 interactions between DENV and humans, as well as 18 interactions between DENV and the *A. aegypti* vector (Doolittle and Gomez 2011). They suggested those specific interactions between virus and host proteins are involved in interferon signaling, transcriptional regulation, stress, and the unfolded protein response.

The most relevant outcome of such computational studies is the identification of human and pathogen proteins to target experimentally for developing new drugs. It also provides different roadmaps and emerging approaches to develop projects to model and analyze PINs of neglected-disease pathogens. For example, novel therapies for human diseases employ multi-target drugs (Borisy *et al.*, 2003, Csermely *et al.*, 2005) and compounds targeted to inhibit protein-protein interactions (Emerson *et al.*, 2003, Klein and Vassilev 2004, Vassilev 2004, Vassilev *et al.*, 2004).

## 5. Conclusions

Because of the development of massive analysis technologies in genomics and computational biology, we can outline a trend to interplay and integrate the computational and experimental techniques. Thus, the methods and resources to identify protein interactions that combine both approaches will be used as a routine protocol in the future.

Even though the use of network biology approaches to drug discovery are in their initial stages, they already contributed to meaningful drug development decisions by accelerating hypothesis-driven biology, modeling specific physiologic problems in target validation or clinical physiology and, providing rapid characterization and interpretation of disease-relevant cell systems.

Despite the lack of experimental functional genomics and PINs data for neglected-disease pathogens, computational approaches represent a starting point and complementary approach to current high-throughput screening projects whose aim is to delineate the complete genomes of neglected-disease pathogens. Moreover, integrative computational approaches have shown to be a powerful tool as guide for large scale-studies improving and facilitating the rational identification of therapeutic targets.

It is clear that for those organisms whose genome has not been sequenced yet, it will be difficult to implement the aforementioned protocols. That is the case for some nematodes and trypanosomal parasites as *T. cruzi*, *S. mansoni*, *B. malayi*, and *O. volvulus*, and the soil-transmitted helminthes (e.g., species of *A. lumbricoides*, and *T. trichura*). However, according to NCBI Entrez Genome (URL: <http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>; Sep 29, 2011), the status of most of them is in "assembly" stage. Once the genome of the neglected-disease pathogen is available, we can use the information of experimental PINs of model organism as *C. elegans* to model and predict PINs of such pathogens enabling the discovery of those hubs and bottlenecks proteins that modulate the infectious process and prioritize them as drug targets.

While the computational approaches analyzed here are by nature probabilistic, i.e. it offers the likelihood of association of a given pair of proteins, nevertheless it clearly indicates the utility of inferring functionally relevant correlations from the available genomic databases for systematic drug target identification. The further improvement of computational approaches will help to increasing the availability of systematically collected biologic data and will provide an easy schema for the integration of different types of data within network analysis, thus enhancing the role of such approaches in drug discovery.

Finally, comprehensive repositories of functional genomic data for neglected-disease pathogens will be created. Hence, as soon as large molecular datasets are processed with the help of network analysis, a growing set of predicted pathways and PINs will emerge and will offer a new paradigm for re-thinking about how to revolutionize the drug discovery process.

## 6. Acknowledgments

The authors thank BioMed Central for allowing the reproduction of figures 8 and 9 (Florez *et al.*, 2010, Navratil *et al.*, 2011). Mario A. Rodríguez-Pérez and Xianwu Guo holds a scholarship from Comisión de Operación y Fomento de Actividades Académicas (COFAA)/IPN.

## 7. References

- Abagyan, R. A., and S. Batalov. (1997). Do aligned sequences share the same fold? *J Mol Biol* 273: 355-368: Oct 17.
- Aguero, F., B. Al-Lazikani, M. Aslett, M. Berriman, F. S. Buckner, R. K. Campbell, S. Carmona, I. M. Carruthers, A. W. Chan, F. Chen, G. J. Crowther, M. A. Doyle, C. Hertz-Fowler, A. L. Hopkins, G. McAllister, S. Nwaka, J. P. Overington, A. Pain, G. V. Paolini, U. Pieper, S. A. Ralph, A. Riechers, D. S. Roos, A. Sali, D. Shanmugam, T. Suzuki, W. C. Van Voorhis, and C. L. Verlinde. (2008). Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov* 7: 900-907: Nov.
- Albert, R. (2005). Scale-free networks in cell biology. *J Cell Sci* 118: 4947-4957: Nov 1.
- Albert, R., and A. L. Barabási. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys* 1: 30.
- Albert, R., H. Jeong, and A. L. Barabasi. (2000). Error and attack tolerance of complex networks. *Nature* 406: 378-382: Jul 27.
- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet* 8: 450-461: Jun.
- Alon, U., M. G. Surette, N. Barkai, and S. Leibler. (1999). Robustness in bacterial chemotaxis. *Nature* 397: 168-171: Jan 14.
- Amaral, L. A., A. Scala, M. Barthelemy, and H. E. Stanley. (2000). Classes of small-world networks. *Proc Natl Acad Sci U S A* 97: 11149-11152: Oct 10.
- Barabasi, A. L., and R. Albert. (1999). Emergence of scaling in random networks. *Science* 286: 509-512: Oct 15.
- Barabasi, A. L., and Z. N. Oltvai. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5: 101-113: Feb.

- Blumenthal, T. (1998). Gene clusters and polycistronic transcription in eukaryotes. *Bioessays* 20: 480-487: Jun.
- Borisy, A. A., P. J. Elliott, N. W. Hurst, M. S. Lee, J. Lehar, E. R. Price, G. Serbedzija, G. R. Zimmermann, M. A. Foley, B. R. Stockwell, and C. T. Keith. (2003). Systematic discovery of multicomponent therapeutics. *Proc Natl Acad Sci U S A* 100: 7977-7982: Jun 24.
- Bowers, P. M., M. Pellegrini, M. J. Thompson, J. Fierro, T. O. Yeates, and D. Eisenberg. (2004). Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol* 5: R35.
- Brenner, S. E., C. Chothia, and T. J. Hubbard. (1998). Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. *Proc Natl Acad Sci U S A* 95: 6073-6078: May 26.
- Buckholz, R. G., C. A. Simmons, J. M. Stuart, and M. P. Weiner. (1999). Automation of yeast two-hybrid screening. *J Mol Microbiol Biotechnol* 1: 135-140: Aug.
- Csermely, P., V. Agoston, and S. Pongor. (2005). The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol Sci* 26: 178-182: Apr.
- Cusick, M. E., N. Klitgord, M. Vidal, and D. E. Hill. (2005). Interactome: gateway into systems biology. *Hum Mol Genet* 14 Spec No. 2: R171-181: Oct 15.
- Chua, H. N., W. K. Sung, and L. Wong. (2007). An efficient strategy for extensive integration of diverse biological data for protein function prediction. *Bioinformatics* 23: 3364-3373: Dec 15.
- Dandekar, T., B. Snel, M. Huynen, and P. Bork. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* 23: 324-328: Sep.
- de Folter, S., R. G. Immink, M. Kieffer, L. Parenicova, S. R. Henz, D. Weigel, M. Busscher, M. Kooiker, L. Colombo, M. M. Kater, B. Davies, and G. C. Angenent. (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. *Plant Cell* 17: 1424-1433: May.
- Deane, C. M., L. Salwinski, I. Xenarios, and D. Eisenberg. (2002). Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics* 1: 349-356: May.
- del Rio, G., D. Koschutski, and G. Coello. (2009). How to identify essential genes from molecular networks? *BMC Syst Biol* 3: 102.
- Devos, D., and A. Valencia. (2001). Intrinsic errors in genome annotation. *Trends Genet* 17: 429-431: Aug.
- Doolittle, J. M., and S. M. Gomez. (2011). Mapping protein interactions between Dengue virus and its human and insect hosts. *PLoS Negl Trop Dis* 5: e954.
- Dunn, R., F. Dudbridge, and C. M. Sanderson. (2005). The use of edge-betweenness clustering to investigate biological function in protein interaction networks. *BMC Bioinformatics* 6: 39.
- Dyer, M. D., T. M. Murali, and B. W. Sobral. (2007). Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics* 23: i159-166: Jul 1.
- Dyer, M. D., T. M. Murali, and B. W. Sobral. (2008). The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog* 4: e32: Feb 8.
- Emerson, S. D., R. Palermo, C. M. Liu, J. W. Tilley, L. Chen, W. Danho, V. S. Madison, D. N. Greeley, G. Ju, and D. C. Fry. (2003). NMR characterization of interleukin-2 in complexes with the IL-2Ralpha receptor component, and with low molecular weight compounds that inhibit the IL-2/IL-Ralpha interaction. *Protein Sci* 12: 811-822: Apr.

- Enright, A. J., and C. A. Ouzounis. (2001). Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions. *Genome Biol* 2: RESEARCH0034.
- Enright, A. J., S. Van Dongen, and C. A. Ouzounis. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575-1584: Apr 1.
- Erdős, P., and A. Rényi. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 4.
- Fields, S., and O. Song. (1989). A novel genetic system to detect protein-protein interactions. *Nature* 340: 245-246: Jul 20.
- Fitzpatrick, D. A., P. O'Gaora, K. P. Byrne, and G. Butler. (2010). Analysis of gene evolution and metabolic pathways using the Candida Gene Order Browser. *BMC Genomics* 11: 290.
- Florez, A. F., D. Park, J. Bhak, B. C. Kim, A. Kuchinsky, J. H. Morris, J. Espinosa, and C. Muskus. (2010). Protein network prediction and topological analysis in Leishmania major as a tool for drug target selection. *BMC Bioinformatics* 11: 484.
- Freeman, L. C. (1977). Set of measures of centrality based on betweenness. *Sociometry* 40: 7.
- Galperin, M. Y., and E. V. Koonin. (1999). Searching for drug targets in microbial genomes. *Curr Opin Biotechnol* 10: 571-578: Dec.
- Galperin, M. Y., and E. V. Koonin. (2000). Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol* 18: 609-613: Jun.
- Galperin, M. Y., and E. V. Koonin. (2004). 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res* 32: 5452-5463.
- Gandhi, T. K., J. Zhong, S. Mathivanan, L. Karthick, K. N. Chandrika, S. S. Mohan, S. Sharma, S. Pinkert, S. Nagaraju, B. Periaswamy, G. Mishra, K. Nandakumar, B. Shen, N. Deshpande, R. Nayak, M. Sarker, J. D. Boeke, G. Parmigiani, J. Schultz, J. S. Bader, and A. Pandey. (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet* 38: 285-293: Mar.
- Gavin, A. C., P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, and G. Superti-Furga. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440: 631-636: Mar 30.
- Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141-147: Jan 10.
- Giot, L., J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, C. E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley, Jr., K. P. White, M.

- Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant, and J. M. Rothberg. (2003). A protein interaction map of *Drosophila melanogaster*. *Science* 302: 1727-1736: Dec 5.
- Girvan, M., and M. E. Newman. (2002). Community structure in social and biological networks. *Proc Natl Acad Sci U S A* 99: 7821-7826: Jun 11.
- Hambly, K., J. Danzer, S. Muskal, and D. A. Debe. (2006). Interrogating the druggable genome with structural informatics. *Mol Divers* 10: 273-281: Aug.
- Hernandez-Toro, J., C. Prieto, and J. De las Rivas. (2007). APID2NET: unified interactome graphic analyzer. *Bioinformatics* 23: 2495-2497: Sep 15.
- Ho, Y., A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthiesen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. Hogue, D. Figeys, and M. Tyers. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415: 180-183: Jan 10.
- Hood, L., and R. M. Perlmutter. (2004). The impact of systems approaches on biological problems in drug discovery. *Nat Biotechnol* 22: 1215-1217: Oct.
- Hopkins, A. L., and C. R. Groom. (2002). The druggable genome. *Nat Rev Drug Discov* 1: 727-730: Sep.
- Huynen, M., B. Snel, W. Lathe, 3rd, and P. Bork. (2000). Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* 10: 1204-1210: Aug.
- Huynen, M. A., and B. Snel. (2000). Gene and context: integrative approaches to genome analysis. *Adv Protein Chem* 54: 345-379.
- Ideker, T., and R. Sharan. (2008). Protein networks in disease. *Genome Res* 18: 644-652: Apr.
- Ito, T., T. Chiba, and M. Yoshida. (2001a). Exploring the protein interactome using comprehensive two-hybrid projects. *Trends Biotechnol* 19: S23-27: Oct.
- Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. (2001b). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98: 4569-4574: Apr 10.
- Ito, T., K. Ota, H. Kubota, Y. Yamaguchi, T. Chiba, K. Sakuraba, and M. Yoshida. (2002). Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol Cell Proteomics* 1: 561-566: Aug.
- Jansen, R., H. Yu, D. Greenbaum, Y. Kluger, N. J. Krogan, S. Chung, A. Emili, M. Snyder, J. F. Greenblatt, and M. Gerstein. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* 302: 449-453: Oct 17.
- Jeong, H., S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. (2001). Lethality and centrality in protein networks. *Nature* 411: 41-42: May 3.
- Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. (2000). The large-scale organization of metabolic networks. *Nature* 407: 651-654: Oct 5.
- Kanehisa, M., and S. Goto. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27-30: Jan 1.

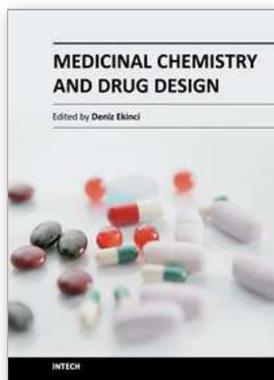
- Kemmeren, P., N. L. van Berkum, J. Vilo, T. Bijma, R. Donders, A. Brazma, and F. C. Holstege. (2002). Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol Cell* 9: 1133-1143: May.
- Klein, C., and L. T. Vassilev. (2004). Targeting the p53-MDM2 interaction to treat cancer. *Br J Cancer* 91: 1415-1419: Oct 18.
- Kohl, P., and D. Noble. (2009). Systems biology and the virtual physiological human. *Mol Syst Biol* 5: 292.
- Koonin, E. V., R. L. Tatusov, and M. Y. Galperin. (1998). Beyond complete genomes: from sequence to structure and function. *Curr Opin Struct Biol* 8: 355-363: Jun.
- LaCount, D. J., M. Vignali, R. Chettier, A. Phansalkar, R. Bell, J. R. Hesselberth, L. W. Schoenfeld, I. Ota, S. Sahasrabudhe, C. Kurschner, S. Fields, and R. E. Hughes. (2005). A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* 438: 103-107: Nov 3.
- Lee, I., S. V. Date, A. T. Adai, and E. M. Marcotte. (2004). A probabilistic functional network of yeast genes. *Science* 306: 1555-1558: Nov 26.
- Lee, I., U. M. Blom, P. I. Wang, J. E. Shim, and E. M. Marcotte. (2011). Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res* 21: 1109-1121: Jul.
- Lee, I., B. Lehner, C. Crombie, W. Wong, A. G. Fraser, and E. M. Marcotte. (2008a). A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nat Genet* 40: 181-188: Feb.
- Lee, I., B. Lehner, T. Vavouri, J. Shin, A. G. Fraser, and E. M. Marcotte. (2010). Predicting genetic modifier loci using functional gene networks. *Genome Res* 20: 1143-1153: Aug.
- Lee, J. H., V. Vittone, E. Diefenbach, A. L. Cunningham, and R. J. Diefenbach. (2008b). Identification of structural protein-protein interactions of herpes simplex virus type 1. *Virology* 378: 347-354: Sep 1.
- Lehner, B., J. I. Semple, S. E. Brown, D. Counsell, R. D. Campbell, and C. M. Sanderson. (2004). Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* 83: 153-167: Jan.
- Li, S., C. M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P. O. Vidalain, J. D. Han, A. Chesneau, T. Hao, D. S. Goldberg, N. Li, M. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari, S. L. Wong, L. V. Zhang, G. F. Berriz, L. Jacotot, P. Vaglio, J. Reboul, T. Hirozane-Kishikawa, Q. Li, H. W. Gabel, A. Elewa, B. Baumgartner, D. J. Rose, H. Yu, S. Bosak, R. Sequerra, A. Fraser, S. E. Mango, W. M. Saxton, S. Strome, S. Van Den Heuvel, F. Piano, J. Vandenhaute, C. Sardet, M. Gerstein, L. Doucette-Stamm, K. C. Gunsalus, J. W. Harper, M. E. Cusick, F. P. Roth, D. E. Hill, and M. Vidal. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* 303: 540-543: Jan 23.
- Linghu, B., E. S. Snitkin, Z. Hu, Y. Xia, and C. Delisi. (2009). Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. *Genome Biol* 10: R91.
- Liu, X., and B. Han. (2009). Evolutionary conservation of neighbouring gene pairs in plants. *Gene* 437: 71-79: May 15.
- Lu, L. J., Y. Xia, A. Paccanaro, H. Yu, and M. Gerstein. (2005). Assessing the limits of genomic data integration for predicting protein networks. *Genome Res* 15: 945-953: Jul.

- Ma'ayan, A., S. L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. J. Eungdamrong, G. Weng, P. T. Ram, J. J. Rice, A. Kershenbaum, G. A. Stolovitzky, R. D. Blitzer, and R. Iyengar. (2005). Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science* 309: 1078-1083: Aug 12.
- Mangan, S., and U. Alon. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* 100: 11980-11985: Oct 14.
- Marcotte, E. M. (2000). Computational genetics: finding protein function by nonhomology methods. *Curr Opin Struct Biol* 10: 359-365: Jun.
- Marcotte, E. M., M. Pellegrini, M. J. Thompson, T. O. Yeates, and D. Eisenberg. (1999a). A combined algorithm for genome-wide prediction of protein function. *Nature* 402: 83-86: Nov 4.
- Marcotte, E. M., M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg. (1999b). Detecting protein function and protein-protein interactions from genome sequences. *Science* 285: 751-753: Jul 30.
- McGary, K. L., T. J. Park, J. O. Woods, H. J. Cha, J. B. Wallingford, and E. M. Marcotte. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci U S A* 107: 6544-6549: Apr 6.
- Mika, S., and B. Rost. (2006). Protein-protein interactions more conserved within species than across species. *PLoS Comput Biol* 2: e79: Jul 21.
- Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. (2002). Network motifs: simple building blocks of complex networks. *Science* 298: 824-827: Oct 25.
- Nair, R., and B. Rost. (2002). Sequence conserved for subcellular localization. *Protein Sci* 11: 2836-2847: Dec.
- Navratil, V., B. de Chasse, C. R. Combe, and V. Lotteau. (2011). When the human viral infectome and disease networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Syst Biol* 5: 13.
- Nikolsky, Y., T. Nikolskaya, and A. Bugrim. (2005). Biological networks and analysis of experimental data in drug discovery. *Drug Discov Today* 10: 653-662: May 1.
- Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 27: 29-34: Jan 1.
- Ouzounis, C., and N. Kyrpides. (1996). The emergence of major cellular processes in evolution. *FEBS Lett* 390: 119-123: Jul 22.
- Overbeek, R., M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev. (1999). The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A* 96: 2896-2901: Mar 16.
- Pedamallu, C. S., and J. Posfai. (2010). Open source tool for prediction of genome wide protein-protein interaction network based on ortholog information. *Source Code Biol Med* 5: 8.
- Pellegrini, M., E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96: 4285-4288: Apr 13.
- Pena-Castillo, L., M. Tasan, C. L. Myers, H. Lee, T. Joshi, C. Zhang, Y. Guan, M. Leone, A. Pagnani, W. K. Kim, C. Krumpelman, W. Tian, G. Obozinski, Y. Qi, S. Mostafavi, G. N. Lin, G. F. Berriz, F. D. Gibbons, G. Lanckriet, J. Qiu, C. Grant, Z. Barutcuoglu, D. P. Hill, D. Warde-Farley, C. Grouios, D. Ray, J. A. Blake, M. Deng, M. I. Jordan, W. S. Noble, Q. Morris, J. Klein-Seetharaman, Z. Bar-Joseph, T. Chen, F. Sun, O. G. Troyanskaya, E. M. Marcotte, D. Xu, T. R. Hughes, and F. P. Roth. (2008). A critical assessment of *Mus musculus* gene function prediction using integrated genomic evidence. *Genome Biol* 9 Suppl 1: S2.

- Prieto, C., and J. De Las Rivas. (2006). APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res* 34: W298-302: Jul 1.
- Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-1555: Aug 30.
- Read, T. D., S. R. Gill, H. Tettelin, and B. A. Dougherty. (2001). Finding drug targets in microbial genomes. *Drug Discov Today* 6: 887-892: Sep 1.
- Rivera, M. C., R. Jain, J. E. Moore, and J. A. Lake. (1998). Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A* 95: 6239-6244: May 26.
- Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng* 12: 85-94: Feb.
- Rozen, R., N. Sathish, Y. Li, and Y. Yuan. (2008). Virion-wide protein interactions of Kaposi's sarcoma-associated herpesvirus. *J Virol* 82: 4742-4750: May.
- Rual, J. F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhoute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437: 1173-1178: Oct 20.
- Russ, A. P., and S. Lampel. (2005). The druggable genome: an update. *Drug Discov Today* 10: 1607-1610: Dec.
- Sakharkar, K. R., M. K. Sakharkar, and V. T. Chow. (2004). A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol* 4: 355-360.
- Sali, A. (1999). Functional links between proteins. *Nature* 402: 23, 25-26: Nov 4.
- Schmid, M. B. (1998). Novel approaches to the discovery of antimicrobial agents. *Curr Opin Chem Biol* 2: 529-534: Aug.
- Schmid, M. B. (2006). Crystallizing new approaches for antimicrobial drug discovery. *Biochem Pharmacol* 71: 1048-1056: Mar 30.
- Shen-Orr, S. S., R. Milo, S. Mangan, and U. Alon. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31: 64-68: May.
- Shoemaker, B. A., and A. R. Panchenko. (2007). Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput Biol* 3: e42: Mar 30.
- Stelzl, U., U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksoz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E. E. Wanker. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122: 957-968: Sep 23.
- Stingl, K., K. Schauer, C. Ecobichon, A. Labigne, P. Lenormand, J. C. Rousselle, A. Namane, and H. de Reuse. (2008). In vivo interactome of *Helicobacter pylori* urease revealed by tandem affinity purification. *Mol Cell Proteomics* 7: 2429-2441: Dec.
- Szklarczyk, D., A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguéz, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen, and C. von Mering. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 39: D561-568: Jan.
- Tamames, J., G. Casari, C. Ouzounis, and A. Valencia. (1997). Conserved clusters of functionally related genes in two bacterial genomes. *J Mol Evol* 44: 66-73: Jan.

- Tew, K. L., X. L. Li, and S. H. Tan. (2007). Functional centrality: detecting lethality of proteins in protein interaction networks. *Genome Inform* 19: 166-177.
- Thanassi, J. A., S. L. Hartman-Neumann, T. J. Dougherty, B. A. Dougherty, and M. J. Pucci. (2002). Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acids Res* 30: 3152-3162: Jul 15.
- Todd, A. E., C. A. Orengo, and J. M. Thornton. (2001). Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* 307: 1113-1143: Apr 6.
- Tong, A. H., G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D. S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J. N. Levinson, H. Lu, P. Menard, C. Munyana, A. B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A. M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S. L. Wong, L. V. Zhang, H. Zhu, C. G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F. P. Roth, G. W. Brown, B. Andrews, H. Bussey, and C. Boone. (2004). Global mapping of the yeast genetic interaction network. *Science* 303: 808-813: Feb 6.
- Uetz, P., and R. E. Hughes. (2000). Systematic and large-scale two-hybrid screens. *Curr Opin Microbiol* 3: 303-308: Jun.
- Uetz, P., Y. A. Dong, C. Zeretzke, C. Atzler, A. Baiker, B. Berger, S. V. Rajagopala, M. Roupelieva, D. Rose, E. Fossum, and J. Haas. (2006). Herpesviral protein networks and their interaction with the human proteome. *Science* 311: 239-242: Jan 13.
- Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403: 623-627: Feb 10.
- Vassilev, L. T. (2004). Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. *Cell Cycle* 3: 419-421: Apr.
- Vassilev, L. T., B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, and E. A. Liu. (2004). In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303: 844-848: Feb 6.
- Venkatesan, K., J. F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K. I. Goh, M. A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J. M. Sahalie, S. Cevik, C. Simon, A. S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M. E. Cusick, F. P. Roth, D. E. Hill, J. Tavernier, E. E. Wanker, A. L. Barabasi, and M. Vidal. (2009). An empirical framework for binary interactome mapping. *Nat Methods* 6: 83-90: Jan.
- Vidal, M., and P. Legrain. (1999). Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* 27: 919-929: Feb 15.
- von Mering, C., R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417: 399-403: May 23.
- Wagner, A., and D. A. Fell. (2001). The small world inside large metabolic networks. *Proc Biol Sci* 268: 1803-1810: Sep 7.
- Walhout, A. J., and M. Vidal. (2001a). High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* 24: 297-306: Jul.

- Walhout, A. J., and M. Vidal. (2001b). Protein interaction maps for model organisms. *Nat Rev Mol Cell Biol* 2: 55-62: Jan.
- Wang, P. I., and E. M. Marcotte. (2010). It's the machine that matters: Predicting gene function and phenotype from protein networks. *J Proteomics* 73: 2277-2289: Oct 10.
- Wang, Y., T. Cui, C. Zhang, M. Yang, Y. Huang, W. Li, L. Zhang, C. Gao, Y. He, Y. Li, F. Huang, J. Zeng, C. Huang, Q. Yang, Y. Tian, C. Zhao, H. Chen, H. Zhang, and Z. G. He. (2010). Global protein-protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. *J Proteome Res* 9: 6665-6677: Dec 3.
- Watts, D. J., and S. H. Strogatz. (1998). Collective dynamics of 'small-world' networks. *Nature* 393: 440-442: Jun 4.
- Weston, A. D., and L. Hood. (2004). Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J Proteome Res* 3: 179-196: Mar-Apr.
- Wiles, A. M., M. Doderer, J. Ruan, T. T. Gu, D. Ravi, B. Blackman, and A. J. Bishop. (2010). Building and analyzing protein interactome networks by cross-species comparisons. *BMC Syst Biol* 4: 36.
- Wu, J., T. Vallenius, K. Ovaska, J. Westermarck, T. P. Makela, and S. Hautaniemi. (2009). Integrated network analysis platform for protein-protein interactions. *Nat Methods* 6: 75-77: Jan.
- Wu, M., X. Li, H. N. Chua, C. K. Kwok, and S. K. Ng. (2010). Integrating diverse biological and computational sources for reliable protein-protein interactions. *BMC Bioinformatics* 11 Suppl 7: S8.
- Wuchty, S., Z. N. Oltvai, and A. L. Barabasi. (2003). Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nat Genet* 35: 176-179: Oct.
- Xia, Y., L. J. Lu, and M. Gerstein. (2006). Integrated prediction of the helical membrane protein interactome in yeast. *J Mol Biol* 357: 339-349: Mar 17.
- Yook, S. H., Z. N. Oltvai, and A. L. Barabasi. (2004). Functional and topological characterization of protein interaction networks. *Proteomics* 4: 928-942: Apr.
- Yu, H., D. Greenbaum, H. Xin Lu, X. Zhu, and M. Gerstein. (2004a). Genomic analysis of essentiality within protein networks. *Trends Genet* 20: 227-231: Jun.
- Yu, H., P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein. (2007). The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 3: e59: Apr 20.
- Yu, H., N. M. Luscombe, H. X. Lu, X. Zhu, Y. Xia, J. D. Han, N. Bertin, S. Chung, M. Vidal, and M. Gerstein. (2004b). Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res* 14: 1107-1118: Jun.
- Yu, H., P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J. F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrikapa, C. Fan, A. S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A. L. Barabasi, J. Tavernier, D. E. Hill, and M. Vidal. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science* 322: 104-110: Oct 3.
- Zorio, D. A., N. N. Cheng, T. Blumenthal, and J. Spieth. (1994). Operons as a common form of chromosomal organization in *C. elegans*. *Nature* 372: 270-272: Nov 17.



## Medicinal Chemistry and Drug Design

Edited by Prof. Deniz Ekinici

ISBN 978-953-51-0513-8

Hard cover, 406 pages

**Publisher** InTech

**Published online** 16, May, 2012

**Published in print edition** May, 2012

Over the recent years, medicinal chemistry has become responsible for explaining interactions of chemical molecules processes such that many scientists in the life sciences from agronomy to medicine are engaged in medicinal research. This book contains an overview focusing on the research area of enzyme inhibitors, molecular aspects of drug metabolism, organic synthesis, prodrug synthesis, in silico studies and chemical compounds used in relevant approaches. The book deals with basic issues and some of the recent developments in medicinal chemistry and drug design. Particular emphasis is devoted to both theoretical and experimental aspect of modern drug design. The primary target audience for the book includes students, researchers, biologists, chemists, chemical engineers and professionals who are interested in associated areas. The textbook is written by international scientists with expertise in chemistry, protein biochemistry, enzymology, molecular biology and genetics many of which are active in biochemical and biomedical research. We hope that the textbook will enhance the knowledge of scientists in the complexities of some medicinal approaches; it will stimulate both professionals and students to dedicate part of their future research in understanding relevant mechanisms and applications of medicinal chemistry and drug design.

### How to reference

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

Aldo Segura-Cabrera, Carlos A. García-Pérez, Mario A. Rodríguez-Pérez, Xianwu Guo, Gildardo Rivera and Virgilio Bocanegra-García (2012). Analysis of Protein Interaction Networks to Prioritize Drug Targets of Neglected-Diseases Pathogens, Medicinal Chemistry and Drug Design, Prof. Deniz Ekinici (Ed.), ISBN: 978-953-51-0513-8, InTech, Available from: <http://www.intechopen.com/books/medicinal-chemistry-and-drug-design/analysis-of-protein-interaction-networks-to-prioritize-drug-targets-of-neglected-diseases-pathogens>

# 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](http://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