Disponible en ligne sur ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

PATHOLOGIE BIOLOGIE

Pathologie Biologie 57 (2009) 324–333

Review

Interaction networks: From protein functions to drug discovery. A review

Les réseaux d'interactions : de la fonction des protéines à la conception de médicaments. Une revue

E. Chautard [a], N. Thierry-Mieg [b], S. Ricard-Blum [a,*]

[a] *UMR 5086 CNRS, institut de biologie et chimie des protéines, université Lyon 1, IFR, 128 biosciences Lyon-Gerland, 7, passage du Vercors, 69367 Lyon cedex 07, France*

[b] *TIMC-IMAG Laboratory, TIMB Team, faculté de médecine, 38706 La Tronche cedex, France*

Received 22 September 2008; accepted 17 October 2008
Available online 13 December 2008

Abstract

Most genes, proteins and other components carry out their functions within a complex network of interactions and a single molecule can affect a wide range of other cell components. A global, integrative, approach has been developed for several years, including protein–protein interaction networks (interactomes). In this review, we describe the high-throughput methods used to identify new interactions and to build large interaction datasets. The minimum information required for reporting a molecular interaction experiment (MIMIx) has been defined as a standard for storing data in publicly available interaction databases. Several examples of interaction networks from molecular machines (proteasome) or organelles (phagosome, mitochondrion) to whole organisms (viruses, bacteria, yeast, fly, and worm) are given and attempts to cover the entire human interaction network are discussed. The methods used to perform the topological analysis of interaction networks and to extract biological information from them are presented. These investigations have provided clues on protein functions, signalling and metabolic pathways, and physiological processes, unraveled the molecular basis of some diseases (cancer, infectious diseases), and will be very useful to identify new therapeutic targets and for drug discovery. A major challenge is now to integrate data from different sources (interactome, transcriptome, phenome, localization) to switch from static to dynamic interaction networks. The merging of a viral interactome and the human interactome has been used to simulate viral infection, paving the way for future studies aiming at providing molecular basis of human diseases.

Résumé

La plupart des gènes, des protéines et des autres constituants cellulaires exercent leurs fonctions au sein d'un réseau complexe d'interactions et une seule molécule peut affecter un ensemble d'autres molécules. Une approche globale, intégrative, a été développée depuis plusieurs années pour construire des réseaux d'interactions protéine–protéine (interactomes). Dans cette revue sont décrites les méthodes dites « haut débit » qui permettent d'identifier plusieurs milliers d'interactions en parallèle. L'information minimale pour décrire une expérience d'interaction moléculaire (MIMIx) a été définie pour le stockage des données dans des bases de données d'interactions publiquement disponibles. Plusieurs exemples de réseaux d'interactions sont donnés, des machines moléculaires (protéasome) ou des organelles (phagosome, mitochondrie) jusqu'aux organismes entiers (virus, bactéries, levure, mouche et vers). Les tentatives de construction de l'interactome humain entier ainsi que les méthodes d'analyse globale des réseaux d'interactions sont brièvement présentées. Ces études ont permis de prédire les fonctions de protéines, d'étudier des voies de signalisation ou métaboliques, de déterminer les mécanismes moléculaires de certaines pathologies (cancer, infections virales) et seront très utiles pour identifier de nouvelles cibles thérapeutiques et pour la mise au point de nouveaux médicaments. Un des défis majeurs est maintenant d'intégrer des données de différentes sources (interactome, transcriptome, phénome, localisome) pour passer des réseaux statiques aux réseaux

* Corresponding author.
E-mail address: s.ricard-blum@ibcp.fr (S. Ricard-Blum).

0369-8114/$ – see front matter
doi:10.1016/j.patbio.2008.10.004

dynamiques. La fusion d'un interactome viral à l'interactome humain a ainsi permis de simuler une infection virale. Cette approche ouvre la voie à des études visant à élucider les bases moléculaires des pathologies humaines.

Keywords: Bioinformatics; Databases; Disease; Drug discovery; Interaction network; Interactome; Protein–protein interactions; Protein–carbohydrate interactions

Mots clés : Bioinformatique ; Bases de données ; Découverte de médicaments ; Interactions protéines–protéines et protéines–sucres ; Rôle des protéines ; Interactome ; Pathologies ; Réseau d'interactions

1. Introduction

Biological systems are made up of very large numbers of different components interacting at various scales. Most genes, proteins and other cell components carry out their functions within a complex network of interactions and a single component can affect a wide range of other components. Interactions involved in biological processes have been first characterized individually, but this ''reductionist'' approach suffers from a lack of information about time, space, and context in which the interactions occur in vivo. A global, integrative, approach has been developed for several years, focusing on the building of protein–protein interaction maps (interactomes). These interaction networks are complex systems, where new properties arise. This is part of an emergent field, called systems biology which is ''the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life'' (http://www.systemsbiology.org/). This interdisciplinary approach, involving techniques from the mathematical, computational, physical and engineering sciences is required to understand complex networks. The systems biology approach has been recently applied to the study of proteases that operate in linear pathways or in regulatory circuits forming a protease web [1,2]. The overexpression or the reduced expression of a protease may perturb the protease web leading either to further connections or to a loss of interactions that may initiate and propagate pathological events [1]. Degradomics, the application of genomic and proteomic approaches to identify the protease and protease-substrate repertoires (degradomes) on an organism-wide scale, has been developed by using specific DNA microarrays to analyze the expression of proteases and inhibitors on a system-wide basis (70-mer oligonucleotide probes for all 1561 human and murine proteases, inactive homologues and inhibitors) and mass spectrometry-based proteomics. Elucidating the substrate degradomes of proteases will help to understand the function of proteases in development and disease and the identification of central proteases will identify new drug targets and will help predicting the potential for side effects due to the interconnected nature of the protease web [2,3]. Systems biology may also be helpful in medicine where treatments focused on components are currently disease-driven, aimed for normalcy and additive, whereas systems biology looks at interrelationships and dynamics. Systems approach will lead to individualized, time-sensitive, space-sensitive and synergistic treatments taking into account the multidimensional use of drugs [4].

An interactome is the whole set of molecular physical interactions between biological entities in cells and organisms and it is essential to understand how gene functions and regulations are integrated at the level of an organism. Indeed, many proteins mediate their biological function through protein interactions, which are involved in supramolecular assemblies (collagens, elastic fibers, actin filaments), in the building of molecular machines (molecular motors, ribosomes, proteasome) and in major biological processes such as immunity (antigen–antibody interaction), metabolism (enzyme–substrate interaction), signalling (interaction of messenger molecules, hormones, neurotransmitters with their cognate receptors), and gene expression (DNA–protein interactions). Furthermore, the sequencing of the genome and advances in proteomics lead to the identification of proteins of unknown functions. Interaction networks might give clues on the functions of these newly discovered proteins or on new functions of already identified proteins. The systematic identification of interactions for a given proteome has been proposed as a potentially informative functional strategy [5,6]. In this review we will:

- describe high-throughput methods used to identify interactions and to build interaction networks, including standard format of reporting and a brief description of publicly available interaction databases;
- give some examples of interaction networks from subcellular compartments to whole organisms (yeast, bacteria, fly, worm);
- explain how to analyze the interaction networks to get information on the biological functions of proteins, and to predict the behavior of the network by simulating constraints induced by physiopathological processes leading to a more rational approach in therapeutics and drug discovery.

2. Building interaction networks

Manual curation and text mining are used to extract interaction data from the literature. New interaction data are collected by high throughput experimental methods, including yeast two hybrid (Y2H), tandem-affinity mass spectrometry (TAP-MS), protein complementation assays and protein arrays. In addition, several methods aiming at predicting interactions (inference, Rosetta stone) have been developed to establish comprehensive maps of hypothetical protein–protein interactions.

2.1. Experimental methods

In yeast two-hybrid, the two proteins to be tested for interactions are expressed in yeast as hybrid fusion proteins

([7]). The principle of the yeast two-hybrid system is the reconstitution of a transcription factor via a protein–protein interaction. The DNA binding domain of the transcription factor is expressed as a hybrid protein fused to protein X (the bait), the activation domain is fused to protein Y (the prey). If the two proteins X and Y interact, the activation domain is in the proper position to activate transcription of the reporter gene. The preys can be individual fusion proteins tested in a one-on-one fashion (the so-called ''matrix'' or ''array'' approach), or they can be a mixture of protein fragments expressed from cDNA libraries from which the interacting preys are selected and identified by sequencing (the ''screen'' approach). Y2H allows high throughput screening of protein interactions and is widely used to build protein–protein interaction networks since being reported by Fields and Song [8]. Further methods have been developed to increase the efficiency and/or coverage of Y2H, such as pooling-sequencing [9] and smart-pooling [10]. The first comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae* by Y2H was published in 2000 by Uetz et al. [11]. They screened nearly all of the 6000 predicted yeast proteins, expressed as DNA-binding domain fusion proteins, against a library, and detected 957 putative interactions involving 1004 *S. cerevisiae* proteins [11]. Another comprehensive two-hybrid analysis to explore the yeast protein interactome identified 4549 interactions among 3278 proteins [12] but the overlap of the two yeast datasets was rather small (141 interactions). A two-hybrid-based protein-interaction map of the fly (*Drosophila melanogaster*) proteome has been built by screening more than 10,000 predicted transcripts against cDNA libraries to produce a draft map of 7048 proteins and 20,405 interactions, which was refined to give a higher confidence map of 4679 proteins and 4780 interactions [13]. Y2H screens have also been used to establish the interactome of *C. elegans* [14] and the first drafts of the human protein–protein interaction network [9,15]. Three thousand one hundred and eighty-six interactions among 1064 baits and 1075 prey proteins were identified in the first dataset [15] and 2754 interactions involving 1549 proteins in the second one [9].

An affinity purification method coupled to mass spectrometry-based protein identification, referred to as tandem affinity purification/mass spectrometry, has been first used at a large scale by Gavin [16], who have identified 1440 proteins within 232 multiprotein complexes in *S. cerevisia*e [17]. TAP-MS involves biochemical isolation of protein complexes formed within yeast cells between tagged proteins (baits) and endogenous yeast proteins and subsequent identification using mass spectrometry TAP-MS has proven successful for retrieval of protein complexes and interacting proteins in *E. coli* [18], the protein interaction network of the TNF-α/NF-κB pathway in human [19], and in yeast [17,20–22]. Gavin et al. [21] have reported the first genome-wide screen for complexes in an organism (*S. cerevisiae*), using TAP-MS and shown that the ensemble of cellular proteins partitions into 491 complexes. Another affinity purification approach referred to as luminescence-based mammalian interactome (LUMIER) has been developed to map protein–protein interaction networks systematically in mammalian cells [23]. This strategy uses *Renilla* luciferase enzyme fused to proteins of interest, which are then co-expressed with individual Flag-tagged partners in mammalian cells. Their interactions are determined by performing a *Renilla* luciferase enzymatic assay on immunoprecipitates using an antibody against Flag. Nine hundred and forty-seven interactions were detected in the TGFβ interaction network using this approach [23].

Protein arrays are used to screen a large number of interactions in parallel. Zhu et al. [24] have constructed a yeast proteome microarray containing approximately 80% yeast proteins (5800 proteins). The proteins were printed onto slides at high spatial density and screened for their ability to interact with proteins and phospholipids [24]. A quantitative protein interaction network for the ErbB receptors has been studied using protein microarrays in experimental conditions that allow the calculation of the apparent equilibrium dissociation constant for every interaction [25]. Both studies used fluorescence as a detection method. We have developed protein and glycosaminoglycan arrays probed by surface plasmon resonance, which does not require the labelling of the interactants and allows the calculation of kinetic and affinity constants of the interactions. We used an automated microarray platform able to monitor up to 400 interactions simultaneously (Biacore Flexchip, GE Healthcare).

Protein-fragment complementation assays have been used to perform a genome-wide in vivo screen for protein–protein interactions [26]. Two proteins of interest are fused to complementary fragments of a reporter protein (the enzyme dihydrofolate reductase) and their physical interaction reconstitutes the activity of the enzyme. This method detects structural and topological relationships between proteins, providing an 8-nanometer resolution map of dynamically interacting complexes in vivo. Two thousand seven hundred and seventy interactions among 1124 endogenous proteins have been identified in yeast, revealing a previously unknown space of the yeast interactome and giving insights into cell polarization and autophagy [26].

2.2. Literature curation: collecting existing knowledge

Interaction data can be retrieved from the literature by hand-curation of papers by biologists. However, extraction of the large amounts of interaction data between proteins from the literature, more than 16 million citations in the Medline database [27], requires mathematical and statistical methods to achieve a comprehensive coverage. The automatic extraction of protein–protein interactions ranges from simple statistical methods relying on co-occurrences of genes or proteins to methods employing syntactical or semantical analysis [27,28]. Current approaches can be divided into computational linguistics-based methods, rule-based methods and machine learning and statistical methods. Several protein-protein interaction information extraction systems and tools for biomedical literature mining are available on the web [27]. Developed and applied natural language processing and literature-mining algorithms have been used to recover 6580 interactions among 3737 human proteins from Medline abstracts [29].

2.3. Computational prediction of protein–protein interactions

These methods are complementary to high-throughput experimental studies. They are based on the structural, genomic, and biological context of proteins and genes in complete genomes for the large-scale prediction of protein–protein interactions, either direct physical interaction between two proteins and/or an indirect functional association of two proteins such as involvement in the same biochemical pathway [30]. Protein interaction maps for complete genomes can be generated using gene fusion events [31,32]. The Rosetta stone method is a computational method based on the analysis of gene fusion patterns of protein domains. It relies upon the fact that some pairs of proteins have homologs in another organism fused into a single protein termed a Rosetta Stone sequence because it deciphers the interaction between the protein pairs [32]. Six thousand eight hundred and nine putative protein–protein interactions have been inferred in *Escherichia coli* and 45,502 in yeast using this method [32]. Kamburov et al. [33] have generated a set of over two million predicted interactions encompassing 696,894 proteins in 184 species or strains using a variant of this method.

Protein–protein interactions can be computationally predicted from co-evolution events. This approach is based on the fact that large numbers of physically interacting proteins in one organism have co-evolved so that their respective orthologs in other organisms (interologs, [34]), interact as well [35]. Interaction maps from one species may indeed be useful in predicting interactions in other species and may provide clues on the function of proteins [35]. Lehner et al. [36] (2004) described a network of over 70,000 predicted physical interactions between around 6200 human proteins generated using the data from lower eukaryotic protein–interaction maps. Inferred human interactions based on orthology mapping of protein interactions discovered in model organisms is stored in the HomoMINT database (http://mint.bio.uniroma2.it/HomoMINT/) [37] and interologous networks of multiple organisms (worm, fly, human, mouse, rat and yeast) are available through the Interolog interaction database (I2D, http://ophid.utoronto.ca/i2d/) [38] that includes 200,599 predicted interactions.

3. Storage of large-set interaction data: interaction databases

Many databases have been generated to store interaction data. They differ by the amount and quality of data, the species involved and the type of interaction (physical and/or functional interactions). Several publicly available databases are discussed below. Interactions stored in IntAct (http://www.ebi.ac.uk/intact) [39] are derived from literature curation or direct user submissions and are freely available. The IntAct Database contains 170,831 binary interactions, 63,824 proteins, and 8827 experiments (August 2008). The molecular interaction database (MINT, http://mint.bio.uniroma2.it) [40] stores 105,899 interactions involving 28,817 proteins and 3647 publications. This database has two sister databases, HomoMINT an inferred human network and Domino, a database of interactions mediated by protein recognition modules. The database of interacting protein (DIP, http://dip.doe-mbi.ucla.edu) [41] stores data that are curated manually by expert curators and automatically using computational approaches. Fifty six thousand six hundred and thirty eight interactions involving 19,935 proteins in 204 organisms corresponding to 64,241 experiments are reported from 3516 articles and 34 other data sources. The biological general repository for interaction datasets (BIOGRID, http://www.thebiogrid.org) is a database of physical and genetic interactions [42]. Other databases are specialized and cover a restricted interaction space. MPACT, the Munich information center for protein sequences (MIPS, http://mips.gsf.de/genre/proj/mpact) database, is a manually annotated protein interaction database in yeast [43]. The microbial protein interaction database (MPIDB, http://www.jcvi.org/mpidb/) aims to collect and provide all known physical microbial interactions. Twenty three thousand five hundred and twenty one experimentally determined interactions among proteins of 193 bacterial species/strains are stored in the database [44]. MatrixDB (http://matrixdb.ibcp.fr, Chautard et al. in revision) is focused on the extracellular matrix. This database currently stores 1433 protein–protein and 109 protein–glycosaminoglycan interactions. Interactions involving multimers (e.g. collagens) and fragments issued from extracellular molecules displaying biological activities of their own (matrikins/matricryptins) are also reported. MatrixDB integrates extracellular interaction data from a general interaction database (human protein reference database, http://www.hprd.org), from manual literature curation and from experiments performed with protein and glycosaminoglycan arrays. Several databases have formed the International molecular interaction exchange (IMEx, http://imex.sf.net) consortium to share the curation load and to regularly interchange data curated to the same common standards. Interactions are reported using the minimum information required for reporting a molecular interaction experiment (MIMIx) [45]. The database of polyanion-binding protein (http://pabp.bcf.ku.edu/DB_PABP/) is focused on interactions of a subset of proteins, the polyanion-binding proteins, with polyanions such as DNA and heparin [46]. Currently, the DB-PABP has 512 proteins involved in 706 PA/PABP interactions retrieved from 205 literature references.

4. Existing interactomes: from subcellular organelles to organisms

A high number of interaction networks has already been built, ranging from molecular machines (26 S proteasome) [47], subcellular organelles such as mitochondrion [48] and phagosome [49], individual cells such as red blood cell [50] to whole organisms such as yeast [11,12,17,20–22], *Drosophila melanogaster* [13], *Caenorhabditis elegans* [14]. Furthermore, a cell interaction knowledgebase, an online

database for innate immune cells, cytokines and chemokines, has been developed recently (http://cell-interaction.bii.a-star.edu.sg/) [51]. The first draft of the extracellular interactome has been recently established (Chautard et al., in revision) and the interaction network of membrane proteins involved in cell adhesion, the integrin adhesome, is comprised of 156 components linked by 690 interactions [52].

4.1. Micro-organisms

The interaction map of several microorganisms have been established to understand their biology, their interference with the cellular functioning, and to identify possible key intervention points for the future development of therapeutics based on their ability to modulate interactions involved in biological functions. Interaction networks have been built for *E. coli* [18,53] with more than 5000 interactions [18], the cyanobacterium *Synechocystis* [54], the nitrogen-fixing bacterium *Mesorhizobium loti* [55], and for a number of bacteria such as the gastric pathogen *Helicobacter pylori* [56], *Campylobacter jejuni*, a major cause of gastroenteritis [57], and *Treponema pallidum*, the syphilis spirochete [58]. Interaction maps of viruses are indeed smaller due to the lower size of their proteome and they have been published for the hepatitis C virus [59], and *vaccinia* virus, which was used as the smallpox vaccine and is currently used as a mammalian expression vector [60]. An analysis of the protein interaction network of the malaria parasite *Plasmodium falciparum* has identified 2846 interactions involving 25% of the predicted *P. falciparum* proteins [61].

4.2. The human interactome

The first draft of human interaction map comprising over 70,000 predicted physical interactions between 6231 human proteins was generated using data from lower eukaryotic protein-interaction [36]. Automated and high throughput yeast two-hybrid experiments identified 3186 interactions among 1705 proteins, resulting in a large, highly connected network [15]. Using topological and Gene Ontology criteria, a scoring system was developed to define 911 high-confidence interactions among 401 proteins. Rual et al. [9] detected ∼2800 human binary protein–protein interactions, revealing more than 300 new connections to over 100 disease-associated proteins. The number of interactions in humans has been estimated to be ∼650,000 using a statistical procedure based on inference [62]. Available human protein interaction data are gathered in the unified human interactome (UniHI) database (http://www.mdc-berlin.de/unihiit) [63]. It is based on 10 major interaction maps derived by computational and experimental methods and includes more than 150,000 distinct interactions between more than 17,000 unique human proteins. The full coverage of the human interactome is complicated by the existence of different cell types and cellular localizations in contrast to yeast. Issues and strategies for the mapping of the human protein interactome are discussed in [64].

5. How to make sense of protein interaction networks: analysis of their topology

The analysis of proteins interaction networks sheds light on the global organization of proteomes but can also place individual proteins into a functional context. In interaction networks, biomolecules are nodes and interactions connecting two nodes (two biomolecules) are edges. Interaction networks are visualized using softwares such as Cytoscape (http://www.cytoscape.org/, Fig. 1). Quantitative description of the networks helps to characterize various biological systems. Network measures that allow the comparison and characterization of networks (degree, degree distribution, shortest path and mean path length, clustering coefficient) and their biological significance have been reviewed [65]. The degree of a node (biomolecule) corresponds to the number of interactions it has with the neighboring molecules. It has been proposed that highly connected proteins (hubs) with a central role in the network architecture are three times more likely to be essential in yeast than proteins with only a small number of interactions [66], but it has been recently shown that protein connectivity correlates with genetic pleiotropy rather than correlating with essentiality [67]. The degree distribution approximates a power law, which indicates that the network is comprised of a few highly connected molecules and that most of the biomolecules have only few links. This is the case in most biological networks, which are called scale-free. They are highly robust to random removal of a protein, but vulnerable to the targeted removal of hub proteins, whose removal drastically alters the network topology [68]. Distance in networks is measured with the path length, which reflects the number of interactions you have to pass through to connect two proteins [65]. The diameter of a network is the maximum distance between two proteins. Other measurements can be computed using freely available softwares such as the centralities in biological networks (CentiBiN) [69].

Two types of computational methods have been developed for network-based prediction of protein function [6]. Direct annotation schemes infer the function of a protein based on its connections in the network because there is a correlation between network distance and functional distance. The closer the two proteins are in the network the more similar are their functional annotations. The second type of approach relies upon module-assisted schemes, which identify modules of related proteins and then annotate each module based on the known functions of its members [6]. Functional modules are composed of many types of molecule. They have discrete functions (protein synthesis, DNA replication, glycolysis) that arise from interactions among their components (e.g. proteins, nucleic acids, polysaccharides). A given molecule may belong to different modules at different times [70]. Some module-assisted functional annotation methods are based solely on topology, whereas others integrate other data such as gene expression or other genomic data [6]. Network analysis tools, a toolbox for the analysis of biological networks, clusters, classes and pathways is freely available on the web (http://rsat.ulb.ac.be/neat/) [71].

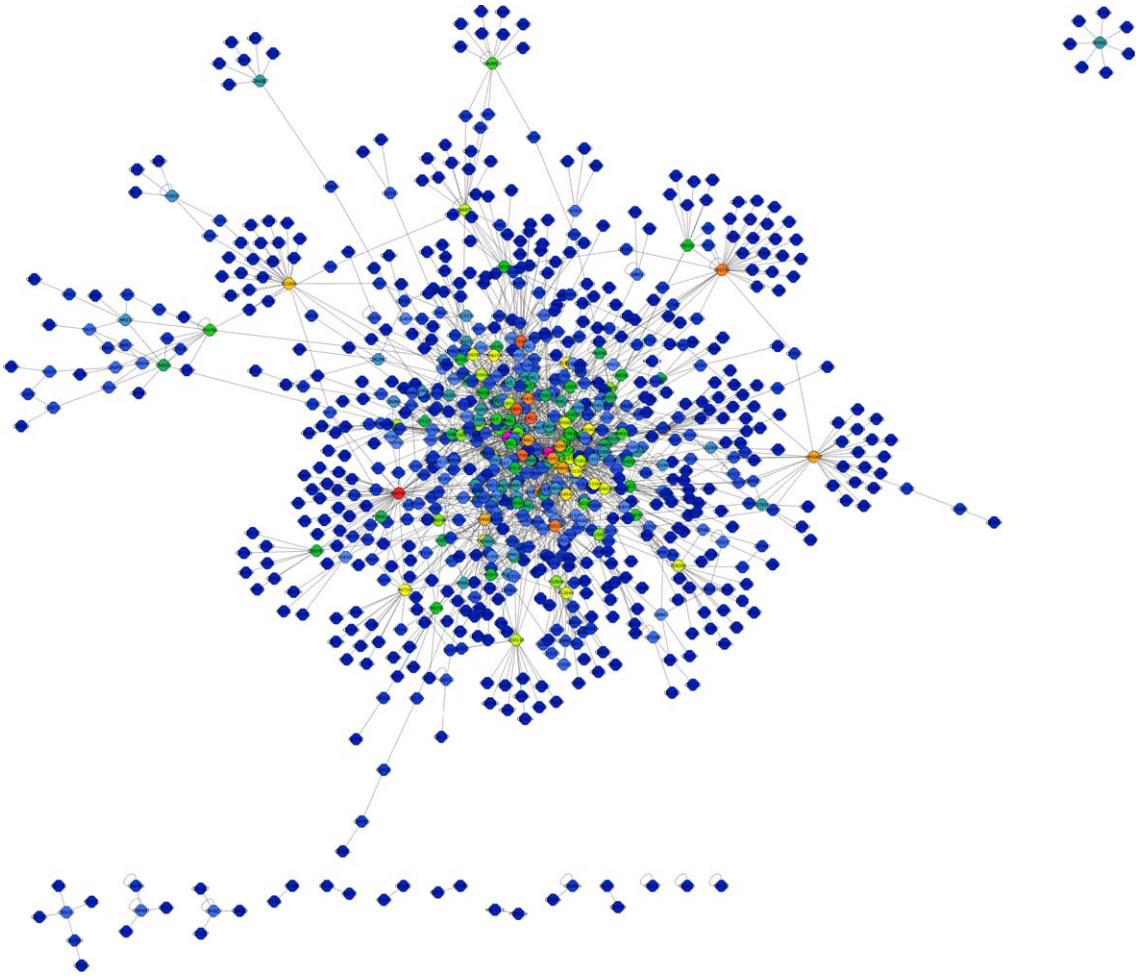

Fig. 1. The extracellular protein–protein and protein–glycosaminoglycan interaction network (859 biomolecules, 1707 interactions). Protein and glycosaminoglycans are represented as squares (nodes), that are color-coded according to the number of their partners (from the lowest to the highest: blue, green, yellow, orange, red and pink). Interactions are represented as links (edges).

6. Some findings resulting from the analysis of protein–protein networks

6.1. Clues on protein functions and pathways

Bioinformatic analysis of the protein interaction network in yeast defined 232 distinct multiprotein complexes and proposed new cellular roles for 344 proteins, including 231 proteins with no previous functional annotation [17]. Analysis of the topological features of cancer proteins in the human interactome has shown that proteins known to be susceptible to mutations leading to cancer show an increase in the number of proteins they interact with [72]. These proteins also appear to participate in central hubs rather than peripheral ones, and participate in networks that form the backbone of the proteome [72]. Functional classification of the proteins participating in nine canonical signalling pathways in the *Drosophila* interactome (wingless, hedgehog, notch, decapentaplegic, sevenless, torso, epidermal growth factor receptor, insulin and toll) identified twelve classes which potentially correspond to twelve functional modules and the participation of ten potential new actors to *Drosophila* signalling was predicted [73]. Each signalling pathway was organized in two to three different signalling modules. The organization of the signalling pathways into different modules may provide the flexibility necessary to the functioning of the same signalling pathway in different spatial, cellular or developmental contexts [73].

6.2. Physiological processes

Aging is associated with many diseases, such as cancer, diabetes, cardiovascular diseases and neurodegenerative disorders and this precludes the investigation of the mechanisms underlying the aging process by focusing on a single gene or a single biochemical pathway. The dynamic modular structure of the protein–protein interaction network during human brain aging has been investigated using a new analytical method, developed to integrate the interactome and the transcriptome [74]. This study of the protein–protein interaction network during aging has identified two modules associated with the cellular proliferation (enriched in development- and differentiation-related genes) to differentiation (enriched in nuclear transport and cell-cycle genes) temporal switch that display opposite aging-related changes in expression [74,75].

Some modular changes might be reversible and genes connecting different modules through protein–protein interactions are more likely to affect aging/longevity. Network simulations further suggest that aging might preferentially attack key regulatory nodes that are important for the network stability [75].

6.3. Unraveling the molecular basis of disease and drug discovery

Protein networks are increasingly serving as tools to unravel the molecular basis of diseases. There are promising applications of protein networks to disease in four major areas: (i) identifying new disease genes, (ii) the study of their network properties, (iii) identifying disease-related subnetworks, and (iv) network-based disease classification [76]. The topological analysis of the human protein–protein interaction network demonstrated that the proteins associated to hereditary disease-genes are characterized by a larger degree, tendency to interact with other disease-genes, more common neighbours and quick communication with each other. An automatic classifier capable of identifying genes more likely to be involved in hereditary disease based on the topological patterns predicted 178 novel disease-genes [77].

6.3.1. Neurodegenerative disorders

Protein–protein interaction networks associated with causative proteins of six neurodegenerative disorders (Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, Huntington's disease, dentatorubropallidoluysian atrophy and prion disease) have been investigated to better understand the molecular pathogenesis of these diseases [78]. They found 19 proteins common to the six diseases, which are mainly involved in the apoptosis and mitogen-activated-protein kinase signalling pathways [78]. Lim et al. [79] described a protein–protein interaction network for inherited human ataxias, a group of diseases characterized by degeneration of cerebellar Purkinje cells. Many ataxia-causing proteins share interacting partners and the majority of the ataxia-causing proteins interact either directly or indirectly. Three proteins involved in RNA binding or splicing represent one of the main hubs in the ataxia network and interact with several different ataxia-causing proteins. This suggests that a subset of inherited ataxias might represent disorders of RNA splicing [79].

6.3.2. Infectious diseases

Intraviral, intrabacterial or intraparasitic interactions are of major interest to decipher the biology of these organisms, but infection triggers dramatic changes in the host and the study of host–pathogen interactions are of major interest to understand the molecular basis of infectious diseases. A virus–host interactome, or virhostome, describes all the ways in which the proteins of a virus interact with those of its host [80]. The study of herpes viral protein networks and their interaction with the human proteome has shown that two herpes viruses, Kaposi sarcoma-associated herpes virus and varicella-zoster virus, shared protein interaction network topologies that were distinct from the cellular networks. Indeed, viral networks should be more resistant to the removal of a highly-connected protein than the human, cellular, network. According to the simulations, infection may result in a change to the viral protein interaction network that renders its topology more similar to that of the host cell [81]. A comprehensive mapping of interactions among Epstein-Barr virus proteins and of interactions of viral proteins with human proteins (virhostome for the Epstein-Barr virus) showed that human proteins targeted by viral proteins were enriched for highly connected proteins in the human interactome (consistent with the hypothesis that hub protein targeting is an efficient mechanism to convert pathways to virus use), and for proteins with relatively short paths relative to all other proteins in the human interactome network [82].

The first comprehensive study of the landscape of human proteins interacting with pathogens integrate 10,477 human-pathogen protein–protein interactions for 190 pathogen strains from seven public databases. Nearly all of the human-pathogen protein–protein interactions (98.3%) involve viral systems [83]. Both viral and bacterial pathogens preferentially interact with two classes of human proteins: proteins with many interacting partners and proteins that lie on many shortest paths (central to many pathways) in the human protein–protein interaction network. Many pathogens target the same processes in the human cell, even if they interact with different proteins, including cell cycle regulation, nuclear transport (import of pathogen proteins into the nucleus in an attempt to subvert the host's DNA replication and transcription machinery), and manipulation of host cellular programs such as apoptosis, immune response and activation of NF-κB pathway [83].

6.4. Therapeutic applications and drug discovery

Most drugs act by binding to specific proteins to modulate their biological activities, with affects biological processes. The search tool for interaction of chemicals (Stitch, http://stitch.embl.de/) database [84] contains protein–chemical interaction data for over 68,000 chemicals, including 2200 drugs, and connects them to 1.5 million genes across 373 genomes. Yildirim et al. [85] generated a bipartite graph composed of US Food and Drug Administration–approved drugs and proteins linked by drug–target binary associations to understand drug targets in the context of cellular and disease networks. The analysis of the network showed that drug targets occupy certain regions in the interactome networks, and their topological signatures are different compared with essential proteins and that most drugs are palliative and do not directly perturb the protein(s) corresponding to the underlying cause of disease [85]. A global map of the large-scale organization of all US approved drugs and associated human therapies has been built, bringing new insights on possible strategies for future drug development [86].

7. Conclusion

The availability of protein–protein interaction networks has significantly increased our understanding of the molecular

mechanisms underlying physiopathological processes. The integration of functional genomic and proteomic data to obtain dynamic networks will further improve the level of confidence of biological hypotheses. A particular challenge in integrating physical and genetic maps is to reconcile the variety of interaction types (i.e. genetic and physical interactions) that are currently available [87]. Interactome, phenome, and transcriptome mapping data were integrated for the *C. elegans* germline [88]. Protein structure information have been combined with protein interaction data to identify residues that form part of an interaction interface, predicting 1428 mutations of the online mendelian inheritance in man database, to be related to an interaction defect [89]. A human phenome–interactome network of protein complexes implicated in genetic disorders has been built [90] and a human protein interolog network has been reconstructed using evolutionary conserved network and computational methods to integrate heterogeneous biological data (subcellular localization, tissue-specificity, cell-cycle stage and domain–domain combination) [91]. Beyond protein–protein interactions, the linkage of all genetic disorders (the human disease phenome) with the complete list of disease genes (the disease genome), results in a global view of the diseasome, the combined set of all known disorder/disease gene associations [92]. The virhostome for the Epstein-Barr virus has been combined with a gene-disease map to generate the Epstein-Barr virus diseasome network, a map of the interconnections between viruses, proteins, genes and diseases [80]. A promising application of these networks is to provide information and to formulate hypotheses on human diseases and therapies (drug discovery and targeting) [76]. Another key issue for future network-based analyses is the dynamics of interaction networks. The current dynactome project (http://dynactome.mshri.on.ca/) aims to analyze the dynamic protein interaction network in normal and malignant cells.

References

[1] Overall CM, Kleifeld O. Towards third generation matrix metalloproteinase inhibitors for cancer therapy. Br J Cancer 2006;94:941–6.

[2] Doucet A, Overall CM. Protease proteomics: Revealing protease *in vivo* functions using systems biology approaches. Mol Aspects Med 2008;29:339–58.

[3] Overall CM, Blobel CP. In search of partners: linking extracellular proteases to substrates. Nat Rev Mol Cell Biol 2007;8:245–57.

[4] Ahn AC, Tewari M, Poon CS, Phillips RS. The clinical applications of a systems approach. PLoS Med 2006;3:e209.

[5] Walhout AJ, Vidal M. Protein interaction maps for model organisms. Nat Rev Mol Cell Biol 2001;2:55–62.

[6] Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Mol Syst Biol 2007;3:88.

[7] Ito T, Ota K, Kubota H, Yamaguchi Y, Chiba T, Sakuraba K, et al. Roles for the two-hybrid system in exploration of the yeast protein interactome. Mol Cell Proteomics 2002;1:561–6.

[8] Fields S, Song O. A novel genetic system to detect protein–protein interactions. Nature 1989;340:245–6.

[9] Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein–protein interaction network. Nature 2005;437:1173–8.

[10] Thierry-Mieg N. A new pooling strategy for high-throughput screening: the shifted transversal design. BMC Bioinformatics 2006;7:28.

[11] Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 2000;403:623–7.

[12] Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 2001;98:4569–74.

[13] Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, et al. A protein interaction map of *Drosophila melanogaster*. Science 2003;302:1727–36.

[14] Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, et al. A map of the interactome network of the metazoan *C. elegans*. Science 2004;303:540–3.

[15] Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell 2005;122:957–68.

[16] Gavin AC, Hopf C. Protein co-membership and biochemical affinity purifications. Drug Discovery Today: Technologies 2006;3:325–30.

[17] Gavin AC, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415:141–7.

[18] Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 2005;433:531–7.

[19] Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol 2004;6:97–105. Erratum in: Nat Cell Biol 2004;6:465.

[20] Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002;415:180–3.

[21] Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature 2006;440:631–6.

[22] Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 2006;440:637–43.

[23] Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. Science 2005;307:1621–5.

[24] Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, et al. Global analysis of protein activities using proteome chips. Science 2001;293:2101–5.

[25] Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 2005;439:168–74.

[26] Tarassov K, Messier V, Landry CR, Radinovic S, Molina MM, Shames I, et al. An *in vivo* map of the yeast protein interactome. Science 2008;320:1465–70.

[27] Zhou D, He Y. Extracting interactions between proteins from the literature. J Biomed Inform 2008;41:393–407.

[28] Bunescu R, Ge R, Kate RJ, Marcotte EM, Mooney RJ, Ramani AK, et al. Comparative experiments on learning information extractors for proteins and their interactions. Artif Intell Med 2005;33:139–55.

[29] Ramani AK, Bunescu RC, Mooney RJ, Marcotte EM. Consolidating the set of known human protein–protein interactions in preparation for large-scale mapping of the human interactome. Genome Biol 2005;6:R40.

[30] Skrabanek L, Saini HK, Bader GD, Enright AJ. Computational prediction of protein–protein interactions. Mol Biotechnol 2008;38:1–17.

[31] Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. Nature 1999;402:86–90.

[32] Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. Science 1999;285:751–3.

[33] Kamburov A, Goldovsky L, Freilich S, Kapazoglou A, Kunin V, Enright AJ, et al. Denoising inferred functional association networks obtained by gene fusion analysis. BMC Genomics 2007;8:460.

[34] Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science 2000;287:116–22.
[35] Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, et al. Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or ''interologs''. Genome Res 2001;11:2120–6.
[36] Lehner B, Fraser AG. A first-draft human protein-interaction map. Genome Biol 2004;5:R63.
[37] Persico M, Ceol A, Gavrila C, Hoffmann R, Florio A, Cesareni G. HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. BMC Bioinformatics 2005;6(Suppl. 4):S21.
[38] Brown KR, Jurisica I. Online predicted human interaction database. Bioinformatics 2005;21:2076–82.
[39] Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, et al. IntAct-Open source resource for molecular interaction data. Nucleic Acids Res 2007;35(Database issue):D561–5.
[40] Chatr-Aryamontri A, Zanzoni A, Ceol A, Cesareni G. Searching the protein interaction space through the MINT database. Methods Mol Biol 2008;484:305–17.
[41] Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The Database of interacting proteins: 2004 update. Nucleic Acids Res 2004;32(Database issue):D449–51.
[42] Breitkreutz BJ, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, et al. The BioGRID Interaction Database: 2008 update. Nucleic Acids Res 2008;36(Database issue):D637–40.
[43] Güldener U, Münsterkötter M, Oesterheld M, Pagel P, Ruepp A, Mewes HW, et al. MPact: the MIPS protein interaction resource on yeast. Nucleic Acids Res 2006;34(Database issue):D436–41.
[44] Goll J, Rajagopala SV, Shiau SC, Wu H, Lamb BT, Uetz P. MPIDB: the microbial protein interaction database. Bioinformatics 2008;24: 1743–4.
[45] Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stümpflen V, et al. The minimum information required for reporting a molecular interaction experiment (MIMIx). Nat Biotechnol 2007;25: 894–8.
[46] Fang J, Dong Y, Salamat-Miller N, Middaugh CR. DB-PABP: a database of polyanion-binding proteins. Nucleic Acids Res 2008;36(Database issue):D303–6.
[47] Davy A, Bello P, Thierry-Mieg N, Vaglio P, Hitti J, Doucette-Stamm L, et al. A protein–protein interaction map of the *Caenorhabditis elegans* 26S proteasome. EMBO Rep 2001;2:821–8.
[48] Perocchi F, Jensen LJ, Gagneur J, Ahting U, von Mering C, Bork P, et al. Assessing systems properties of yeast mitochondria through an interaction map of the organelle. PLoS Genet 2006;2:e170.
[49] Stuart LM, Boulais J, Charriere GM, Hennessy EJ, Brunet S, Jutras I, et al. A systems biology analysis of the *Drosophila* phagosome. 2007;445: 95–101.
[50] Goodman SR, Kurdia A, Ammann L, Kakhniashvili D, Daescu O. The human red blood cell proteome and interactome. Exp Biol Med (Maywood) 2007;232:1391–408.
[51] Chen S, Chinnaswamy A, Biswas SK, Goryachev AB, Yap CK, Lam KY, et al. Cell interaction knowledgebase: an online database for innate immune cells, cytokines and chemokines. In Silico Biol 2007;7:569–74.
[52] Zaidel-Bar R, Itzkovitz S, Ma'ayan A, Iyengar R, Geiger B. Functional atlas of the integrin adhesome. Nat Cell Biol 2007;9:858–67.
[53] Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, et al. Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. Genome Res 2006;16:686–91.
[54] Sato S, Shimoda Y, Muraki A, Kohara M, Nakamura Y, Tabata S. A large-scale protein protein interaction analysis in *Synechocystis sp*. PCC6803. DNA Res 2007;14:207–16.
[55] Shimoda Y, Shinpo S, Kohara M, Nakamura Y, Tabata S, Sato S. A large scale analysis of protein-protein interactions in the nitrogen-fixing bacterium *Mesorhizobium loti*. DNA Res 2008;15:13–23.
[56] Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, et al. The protein–protein interaction map of *Helicobacter pylori*. Nature 2001;409:211–5. Erratum in: Nature 2001;409:553. Nature 2001;409: 743.
[57] Parrish JR, Yu J, Liu G, Hines JA, Chan JE, Mangiola BA, et al. A proteome-wide protein interaction map for *Campylobacter jejuni*. Genome Biol 2007;8:R130.
[58] Titz B, Rajagopala SV, Goll J, Häuser R, McKevitt MT, Palzkill T, et al. The binary protein interactome of *Treponema pallidum*-the syphilis spirochete. PLoS One 2008;3:e2292.
[59] Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspé G, Tiollais P, et al. A genomic approach of the hepatitis C virus generates a protein interaction map. Gene 2000;242:369–79.
[60] McCraith S, Holtzman T, Moss B, Fields S. Genome-wide analysis of vaccinia virus protein–protein interactions. Proc Natl Acad Sci USA 2000;97:4879–84.
[61] LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, et al. A protein interaction network of the malaria parasite *Plasmodium falciparum*. Nature 2005;438:103–7.
[62] Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, et al. Estimating the size of the human interactome. Proc Natl Acad Sci USA 2008;105:6959–64.
[63] Chaurasia G, Iqbal Y, Hänig C, Herzel H, Wanker EE, Futschik ME. UniHI: an entry gate to the human protein interactome. Nucleic Acids Res 2007;35(Database issue):D590–4.
[64] Figeys D. Mapping the human protein interactome. Cell Res 2008; 18:716–24.
[65] Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5:101–13.
[66] Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–2.
[67] Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, et al. M. High-quality binary protein interaction map of the yeast interactome network. Science 2008;322:104–10.
[68] Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. Nature 2000;406:378–82.
[69] Junker BH, Koschützki D, Schreiber F. Exploration of biological network centralities with CentiBiN. BMC Bioinformatics 2006;7: 219.
[70] Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature 1999;402:C47–52.
[71] Brohée S, Faust K, Lima-Mendez G, Sand O, Janky R, Vanderstocken G, et al. NeAT: a toolbox for the analysis of biological networks, clusters, classes and pathways. Nucleic Acids Res 2008;36(Web Server issue): W444–51.
[72] Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics 2006;22:2291–7.
[73] Baudot A, Angelelli JB, Guénoche A, Jacq B, Brun C. Defining a modular signalling network from the fly interactome. BMC Syst Biol 2008;2:45.
[74] Xia K, Xue H, Dong D, Zhu S, Wang J, Zhang Q, et al. Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms. PLoS Comput Biol 2006;2:e145.
[75] Xue H, Xian B, Dong D, Xia K, Zhu S, Zhang Z, et al. A modular network model of aging. Mol Syst Biol 2007;3:147.
[76] Ideker T, Sharan R. Protein networks in disease. Genome Res 2008;18:644–52.
[77] Xu J, Li Y. Discovering disease-genes by topological features in human protein-protein interaction network. Bioinformatics 2006;22: 2800–5.
[78] Limviphuvadh V, Tanaka S, Goto S, Ueda K, Kanehisa M. The commonality of protein interaction networks determined in neurodegenerative disorders (NDDs). Bioinformatics 2007;23:2129–38.
[79] Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, et al. A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 2006;125:801–14.
[80] Gulbahce N, Barabási A-L. Viral networks, connecting virus-human interaction networks to population disease patterns. International Workshop and Conference on Network Science (NetSci'08) 2008 June 23–27; Norwich (UK).

[81] Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, et al. Herpesviral protein networks and their interaction with the human proteome. Science 2006;311:239–42.
[82] Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, et al. Epstein-Barr virus and virus human protein interaction maps. Proc Natl Acad Sci USA 2007;104:7606–11.
[83] Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog 2008; 4:e32.
[84] Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 2008;36(Database issue):D684–8.
[85] Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. Nat Biotechnol 2007;25:1119–26.
[86] Nacher JC, Schwartz JM. A global view of drug-therapy interactions. BMC Pharmacol 2008;8:5.
[87] Beyer A, Bandyopadhyay S, Ideker T. Integrating physical and genetic maps: from genomes to interaction networks. Nat Rev Genet 2007;8:699–710.
[88] Walhout AJ, Reboul J, Shtanko O, Bertin N, Vaglio P, Ge H, et al. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. Curr Biol 2002;12:1952–8.
[89] Schuster-Böckler B, Bateman A. Protein interactions in human genetic diseases. Genome Biol 2008;9:R9.
[90] Lage K, Karlberg EO, Størling ZM, Olason PI, Pedersen AG, Rigina O, et al. A human phenome–interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 2007;25:309–16.
[91] Huang TW, Lin CY, Kao CY. Reconstruction of human protein interolog network using evolutionary conserved network. BMC Bioinformatics 2007;8:152.
[92] Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proc Natl Acad Sci USA 2007;104:8685–90.