International Journal of
Analytical, Pharmaceutical and Biomedical Sciences

IJAPBS

Research Article

Volume: 4: Issue-1: January-2015 *Copyrights@2014* *ISSN:2278-0246*

Coden : IJAPBS *www.ijapbs.com*

PROTEIN-PROTEIN INTERACTION DATABASES: AN OVERALL VIEW ON INTERACTOME ORGANIZATION

Nahid Safari-Alighiarloo[1], Mohammad Taghizadeh[2], Mostafa Rezaei-Tavirani[3]

[1]Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: N.Safari@sbmu.ac.ir
[2]Bioinformatics Department, Institute of Biochemistry and Biophysics, Tehran University, Tehran, Iran. Corresponding author; Email: mtaghizadeh@ut.ac.ir
[3]Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: tavirany@yahoo.com

ABSTRACT: Protein-protein interaction (PPI) maps or "interactome" represent a global view of cellular functions and biological processes. There are more than 130 public databases which have been already established to store the sizeable amount of PPI data. The valuable data is available for a wide group of researchers via adopting common data standards and coordinated curation practices between the databases. Three major types of databases have deposited PPI data; primary databases capture experimentally proven protein-interactions from literatures according to different curation polices; metadatabases integrate the work of other databases; prediction databases combine both predicted and experimental PPIs data from different approaches. Several lines of evidence indicate that primary databases report PPI data according to their different organizing strategies; as a result, this issue may affect the structure of derived PPI networks. Since PPI network's structure is useful to get biologically meaningful information, quality of PPI data needs more attention. In this sense, databases should control the quality of data flow and provide more reliable data with transparent annotations.
Keywords: Protein-Protein Interaction, Overall View Interactome Organization.

THE NATURE OF PROTEIN-PROTEIN INTERACTIONS DATA

Proteins control and carry out most cellular process by interact with each other; as, the identification of proteins' interactions is a vital step toward revealing the intricate molecular relationships within cells [1-3]. Protein-protein interaction (PPI) maps or "interactome" improved our understanding about the functional organization of proteome [4-6]. Currently, systemic approaches are being used to generate genome-scale protein interaction maps as high-throughput, which explore two distinct types of protein interactions; binary interactions and multi-protein complexes [7, 8].

Three major techniques have been carried out to detect binary interactions: yeast two hybrid (Y2H) screens, split ubiquitin and protein complementation assay (PCA). The Y2H is an in vivo technique, which detects transient and unstable interactions in the yeast nucleus. However, only two proteins are tested at a time (no cooperative binding). In addition to this, since protein-interactions occur in the nucleus, many proteins are not in their native compartment [9, 10]. To reduce some serious drawbacks of Y2H, other variants of Y2H were designed like Split- Ubiquitin [11] which detects interactions in cellular compartments in addition to the nucleus. This method was applied to determine interactions between membrane proteins in yeast [12]. PCA [13] explore the co-localization of two proteins in the cell as it used to detect in vivo protein interactions as well as their modulation or spatial and temporal changes [14].

A different approach is used for detection of genome-scale multi-protein complexes, involving the coupling of affinity purification methods with mass spectrometry (AP-MS). In an AP-MS experiment, a tagged protein is expressed in in vivo environment (in the cell) and purified from cell extract using the tag, for instance with

specific antibody of this tag. Binding proteins to the tagged protein are co-purified and afterwards identified by MS [15]. There are two protocols: tandem affinity purification (TAP) by which the interest protein is attached a larger protein tag which facilities sequential affinity purification steps and high-throughput mass-spectrometric protein complex identification (HMSPCI) [5, 16, 17]. AP-MS based approaches are an appropriate to detect real complexes in physiological settings; however, some complexes which are not present under the given conditions might be missed. Additionally, the process of tagging may disturb complex formation, and components which don't have tightly association may be washed off during purification [10].

There is a need to organize such massive amount of collected PPI data in a structured format, which helps users to understand the data, to import them in analysis and visualization tools and to compare them with similar data in related resources [18]. Various databases also have been established in order to organize and process this data.

AN OVERVIEW ON STANDARDIZED FORMAT OF PROTEIN INTERACTIONS DATA AND IMEX CONSORTIUM

Sizable amount of protein-interactions have been identified and stored in public databases during the last decade, and the rate of their production is growing exponentially [19]. Indeed, more than 130 resources have been already established including PPI data [20], even if the vast majority of interactions are stored in only a few of them.

The first molecular interaction databases were independently created and had their own database model, curation strategy and data download formats [18]. In 2002, by the efforts of the Human Proteome Organization Proteomics Standards Initiative (HUPO-PSI), the issue of the individual data resource formats of the separate resources has largely been addressed [21]. Two formats of Proteomics Standards Initiative- Molecular Interaction (PSI-MI) for the first time developed as common data formats. Both XML (PSI-MI XML) and Excel-compatible, tab-delimited (MITAB) are independent means of transferring information between differing locations [22, 23]. Additionally, a common controlled vocabulary (CV) was developed to describe each experimental technique, molecular feature and interaction type in a single common word or phrase and also by a computer-parsable accession number [23]. It has achieved a greater consistency of annotation across databases by using this CV within each database or mapping on the production of export files. Further to this, it has facilitated searching for, and filtering out of, selected data for the users [18]. After a while, in 2007, the 'minimum information about a molecular interaction experiment' (MIMIX) guidelines had been prepared for authors to concern a list of the essential information for describing experimental molecular-interaction data in a journal publication [24].

Use of the same data format by multiple resources has salient advantage; such that it provides possibility to simultaneously access multiple resources with a single query. A common query interface (PSICQUIC) web service which provides access to this goal is available by over 28 sources databases [25, 26].

Although a common data format facilitates to accesses consistent, user-friendly publicly available molecular interaction data, it is only a first stage. Until recently, all interaction databases independently performed literature curation. Sometimes, several different datasets derived from a single publication because of the implementation of different curation strategies [27].

To address this challenge, five molecular interaction databases have created the International Molecular Exchange (IMEx) consortium in September 2005 to coordinate their curation strategies [27]. Database of Interacting Proteins (DIP) [28], the IntAct molecular interaction database (IntAct) [29], Molecular Interaction database (MINT) [30] , the extracellular Matrix interaction Database (MatrixDB) [31], Microbial Protein Interaction database (MPIDB) [32], Interologous Interaction Database (I2D) [33], Innate immune response Database (InnateDB) [34] and Molecular Connections (http://www.molecularconnections.com) are as full members and Biological General Repository for Interaction Datasets (BioGRID) [35] is an observer member.

The aim of the IMEx consortium is to share curation efforts across all its member databases in order to maximize curation coverage and synchronize curation strategies. All aspects of an interaction experiment are captured in a deep curation model of this consortium such as full description of any constructs and the tags, mutations or labels that may be associated with them [27]. The enormous space of estimated human interctome which comprises nearly 130,000 – 650,000 protein interactions [36, 37] and an inherently dynamic entity of interactome that varies among tissues, disparate cell process and environmental conditions exacerbate the need to define and enforce standards on the data formats to easily gather new PPIs information [38]. It is to be hoped to get more reliable and complete PPI data to obtain more informative biological interpretations.

MAJOR CATEGORIES OF PPI DATABASES

Overall, there are three major categories of databases that PPI data have been deposited on them. To start with, primary databases capture experimentally proven protein-interactions from literatures which include either small-scale or large-scale published studies. In addition to, there are now several metadatabases which do not themselves annotate PPI data from the literatures but merge the work of other databases using a variety of methodologies. Prediction databases are another major one, which exploiting a variety of information sources such as predicted PPIs from different approaches combined with experimentally verified PPIs [39]. Here, some major PPI databases have been described, and additional resources are founded at http://www.pathguide.org [20]. Description of PPI databases is presented in Table 1.

PRIMARY DATABASES

The major primary databases are Human Protein Reference Database (HPRD) [40], the Biomolecular Interaction Network Database (BIND) [41], DIP, IntAct, MINT and BioGrid which report only experimentally verified interactions. DIP, MINT and IntAct are active members of IMEx consortium, and BioGrid is an observer member. HPRD has not followed the consortium standards, but stays as largest human repositories, and BIND include informative piece of data that have been only deposited on this database; however, it is no longer active. BIND, DIP, HPRD, IntAct and MINT did not incorporate data from other databases. BioGRID imported a noticeable part of the HPRD PPI data at the time of its creation, but it have not added any more data from other databases since then [42]. BioGRID include both protein and genetic interactions from sixty organisms [35]. Among the mentioned databases, HPRD is the one that focuses entirely on human proteins [40]. Furthermore, there are other groups of primary databases that they combined selected data from curated resources with their own curation efforts [27]. MatrixDB (extracellular PPIs), InnateDB (PPIs in the immune system) and MPIDB (PPIs in microbes) are examples of these databases, which provided informative databases for interested in this area.

METADATABASES

Metadatabases unify the PPI data of their selected databases into a single comprehensive source. Agile Protein Interaction DataAnalyzer (APID) extracts interactions from the six experimentally deposited databases: BioGRID, BIND, DIP, HPRD, IntAct and MINT, in which all proteins mapped to UniProt identifiers [43]. The other resource is interaction Reference Index (iRefIndex) [44] which incorporates protein-interaction data from nine interaction databases including BIND, BioGRID, DIP, HPRD, IntAct, MINT, MPact [45], MPPI [46] and OPHID [47]. Unified Human Interactome (UniHI) provides human PPI maps which have been derived from both experimentally and computationally predicted PPI. BioGRID, BIND, DIP, HPRD and IntAct databases are the used databases, and additional data derived from computational approaches such as large Y2H screenings [4, 48], text-mining [49], an orthology based predictions, (OPHID, HomoMint [50] and ORTHO [51]). Michigan Molecular Interactions (MiMI) merges data from BioGRID, BIND, DIP, HPRD, IntAct and MINT, and it also adds interactions from KEGG [52] and Reactome [53] pathway databases together with published databases from Max Delbruck Center [4] and the WSU Campylobacter Interactome [54]. Human Integrated Protein-Protein Interaction rEference (HIPPIE) integrates interactions data from the following public databases: BioGRID, DIP, HPRD, IntAc, MINT, BIND and MIPS. Additionally, interactions from manually selected studies have been merged in that database. High-quality INTeractomes' (HINT) interactions for the organisms were retrieved from the public databases: BioGrid, DIP, HPRD, IntAct, iRefWeb [55], MINT, MIPS [56] and VisAnt [57]. PINA is a network analysis platform, which includes data from MINT, IntAct, DIP, BioGRID, HPRD and MIPS/MPACT. It is also possible to upload user's private PPI data sets on PINA to create meta-databases [58].
ConsensusPath [59] and PathwayCommons [60] are metadabases which integrate the interaction networks into reaction pathway. These databases integrate data from pathway databases with the available information from the interaction databases to create the complex pathway.

PREDICTION DATABASES

STRING, OPHID, PIPs and STITCH are major prediction PPI databases. Search Tool for the Retrieval of Interacting Gene (STRING) integrates experimentally proven protein- interaction data from databases BIND, BioGRID, DIP, IntAct MINT and HPRD with interactions from the pathway databases PID [61], Reactome,

KEGG and EcoCyc [62]. In addition to, this database also includes interactions predicted by algorithms especially made for STRING [63, 64].The Online Predicted Human Interaction Database (OPHID) provides predicted interactions between human proteins accessible on web-based. It incorporates the literature-derived human protein-interactions from three major public databases, BIND, HPRD and MINT, with predictions made from Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster and Mus musculus [47]. The human protein–protein interaction prediction database (PIPs) is a resource of human protein-interactions predicted by a naive Bayesian model as described in Scott and Barton [65]. In short, the method [65]integrates information from gene co-expression, orthology, co-occurrence of domains, post-translational modifications, co-localization of the proteins within the cell and analysis of the local topology of the predicted PPI network. STITCH is an interaction network database for small molecules and proteins. It uses STRING database to find its predicted interactions. It is novel features for STITCH which provides predictions regarding protein–chemical interactions [66].

Table 1: Description of PPI databases

Database	URL	Species	Types of interactions
Primary databases			
DIP	http://dip.doe-mbi.ucla.edu/dip/	All	Only PPI
MINT	http://mint.bio.uniroma2.it/mint/	All	Only PPI
HPRD	http://www.hprd.org/	Human	Only PPI
IntAct	http://www.ebi.ac.uk/intact/	All	PPI @ others
BIND	http://bond.unleashedinformatics.com/	All	PPI @ others
BioGRID	http://www.thebiogrid.org/	All	PPI @ others
Metadatabases			
iRefIndex	http://www.iRefIndex.org/	All	Only PPI
APID	http://bioinfow.dep.usal.es/apid/	All	Only PPI
MiMI	http://mimi.ncibi.org/MimiWeb/	All	PPI @ others
UniHI	http://www.mdc-berlin.de/unihi/	Human	PPI @ others
HIPPIE	http://cbdm.mdc-berlin.de/tools/hippie	Human	Only PPI
HINT	http://hint.yulab.org	Four species	Only PPI
PINA	http://cbg.garvan.unsw.edu.au/pina/	All	Only PPI
Predicted databases			
STRING	http://string.embl.de/	All	PPI @ others

OPHID	http://ophid.utoronto.ca	Human	PPI @ others
PIPs	http://www.compbio.dundee.ac.uk/www-pips/	Human	PPI @ others
STITCH	http://stitch.embl.de/	All	Protein @ small molecule

MORE DETAIL ABOUT MANUAL CURATION OF PRIMARY DATABASES

Primary databases have different number of proteins and interactions (Table 2). Among mentioned databases, two most comprehensive databases in terms of individual interactions are BioGrid and IntAct, with about 563341 (non-redundant) and 293,000 interactions from up to 60 and 1082 different organisms, respectively. HPRD have been restricted to human proteins; it reports about 41,000 unique interactions, whereas BioGRID and IntAct reports about 193,000 and 102,000 unique human protein-protein interactions, respectively.

Table 2: General features of primary PPI databases

Database	Number of proteins	Number of Interactions	Number of Organisms	Number of Experiments	Number of human proteins	Number of human Interactions
DIP	26743	77514	665	74901	4110	6794
MINT	35553	241458	362	16305	8762	26830
HPRD	9673	41,327	1	20,164	9673	41,327
IntAct	82947	292211	1082	32922	15676	101807
*BIND	66754 (33177)	183495 (60770)	1588	unknown	8269 (5347)	15739 (9593)
BioGRID	56901	563341	60	44686	20073	192513

* The statistics are from BIND translate project and iRefweb

There are some points which elucidate discrepancies of the size and entries of these PPI databases. Most of these resources are independently run and funded, and have different curation practices [27]. Indeed, different databases have not followed the same rules for capturing the data. Besides, different systems for cross-referencing genes and proteins across biological databases have been applied in each database. Therefore, PPI data are reported by two different databases from curation of the same publication may have significant differences [67].

The mapping of ambiguous protein descriptions presented in the text to identifiers in sequence databases is the first impediment to merge data of different databases. One problem which further makes more complex the assigning and cross referencing gene and protein identifiers is the annotation of protein isoforms. This information is rarely prepared in original articles. There are several ways by which databases deal with this ambiguity. Addressing this issue done either by mapping the data to a gene identifier and missing a specific isoform (BioGRID) or by selecting one transcript, usually the longest (BIND), which makes it impossible to determine when this is an ambiguous or a specific mapping. Another strategy is to use the canonical sequence (IntAct, MINT, DIP, MatrixDB, I2D and MPIDB) [27] provided by UniProtKB [25].

Another reason of differences between primary databases is related to this point that each database has various policies to describe mentioned interactions between protein constructs from different species such as human and mouse. Most databases report the exact protein species data extracted from the experiment while others map

these interactions onto a single organism such as human (HPRD) [27]. In addition to, some databases' policy is to extract only subset protein-interactions that relates to their specific area of interest; therefore, they may only partially curate a publication to this aim (InnateDB, HPRD and MPIDB) whereas others systematically curate a publication (DIP, MINT, and IntAct) [27]. Although none of these policies are in any way wrong, they make difficulties to conciliate redundancies between databases. One report suggested that the agreement between two databases was about 42% for curated interactions and 62% in protein identifications when they used the same publication as a curation resource. Some difficulties have been mentioned for this matter consist of divergent organism assignments, the use of alternative protein identifiers and different representations of complexes [67]. To handle these challenges, IMEx consortium, develop common curation and data presentation practices, and deliver unique set of protein interactions on a single web portal (http://www. imexconsortium.org).

Considering that, there are huge amount of protein-interactions data, and no PPI databases could cover all the reported interactions. Therefore, the integration of PPI data from various databases seems necessary to increase the chance of obtaining more significant biological interpretations. It should be taken into account as new challenges which aim to resolve the differences among the PPI databases, including the identifiers, confidence and context [27, 68-71]. For instance, when talked about the interaction of two proteins, three points come to mind: the performing of the experiment with constructs of the same species, using of the full-length proteins or variants, and detecting of which cells and r what conditions [71]. In order to perform meaningfully interactome analysis, annotation policies should be comprehensive; such that the user understands the entities of data and their detailed information. Indeed, it is responsibility of the databases to provide the high quality, consistently curated data [18].

Users should take careful attention to some points such as the source of the data in the different databases, their strategies to quality controlling and scoring the underlying data. [72]. Whereas network analysis of PPI data are used in biomedical research, there are other caveats need to be overcome, such as interactome biases, orphans, and the accuracy and context of reported interactions [71].

Some databases are a mix of protein–protein interactions, regulatory associations, synthetic lethality relationships, pathway co-occurrence, and text-mining associations. Although these databases include many types of functional association data in the network to increase coverage, it may be the cost of a more ambiguous interpretation of what the network means from a biochemical and molecular biology viewpoint [72]. It seems that there are the number of databases which only apply PPI data to establish networks like, InWeb [73], IMEx consortium, mentha [74] , I2D [75] and PINA.

Various properties of PPI networks have been used to get biological insights. Network topology has a prominent role in understanding network architecture and performance. Indeed, the organizational principles of biological systems have been described by analyzing of topological properties of PPI networks [76, 77]. In line with these concerns, to get meaningful biological interoperations, PPI data should have highest quality, and quality control of PPI data should be an essential part of any database establishment.

REFERENCES

[1] Alberts, B., The cell as a collection of protein machines: preparing the next genera-tion of molecular biologists. Cell, 1998. 92(3): p. 291-294.
[2] Pellegrini, M., D. Haynor, and J.M. Johnson, Protein interaction networks. Expert review of proteomics, 2004. 1(2): p. 239-249.
[3] Butland, G., et al., Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature, 2005. 433(7025): p. 531-537.
[4] Stelzl, U., et al., A human protein-protein interaction network: a resource for anno-tating the proteome. Cell, 2005. 122(6): p. 957-968.
[5] Gavin, A.-C., et al., Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 2002. 415(6868): p. 141-147.
[6] Yu, H., et al., The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS computational biology, 2007. 3(4): p. e59.
[7] Fields, S. and O.-k. Song, A novel genetic system to detect protein protein interac-tions. 1989.

[8] Gavin, A.-C., K. Maeda, and S. Kühner, Recent advances in charting protein–protein interaction: mass spectrometry-based approaches. Current opinion in bio-technology, 2011. 22(1): p. 42-49.
[9] Ito, T., et al., A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences, 2001. 98(8): p. 4569-4574.
[10] Von Mering, C., et al., Comparative assessment of large-scale data sets of protein–protein interactions. Nature, 2002. 417(6887): p. 399-403.
[11] Thaminy, S., J. Miller, and I. Stagljar, The split-ubiquitin membrane-based yeast two-hybrid system, in Protein-Protein Interactions. 2004, Springer. p. 297-312.
[12] Miller, J.P., et al., Large-scale identification of yeast integral membrane protein interactions. Proceedings of the National Academy of Sciences of the United States of America, 2005. 102(34): p. 12123-12128.
[13] Morell, M., S. Ventura, and F.X. Avilés, Protein complementation assays: ap-proaches for the in vivo analysis of protein interactions. FEBS letters, 2009. 583(11): p. 1684-1691.
[14] Tarassov, K., et al., An in vivo map of the yeast protein interactome. Science, 2008. 320(5882): p. 1465-1470.
[15] Rigaut, G., et al., A generic protein purification method for protein complex charac-terization and proteome exploration. Nature biotechnology, 1999. 17(10): p. 1030-1032.
[16] Puig, O., et al., The tandem affinity purification (TAP) method: a general proce-dure of protein complex purification. Methods, 2001. 24(3): p. 218-229.
[17] Ho, Y., et al., Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature, 2002. 415(6868): p. 180-183.
[18] Orchard, S., Molecular interaction databases. Proteomics, 2012. 12(10): p. 1656-1662.
[19] Ceol, A., et al., Linking entries in protein interaction database to structured text: the FEBS Letters experiment. FEBS letters, 2008. 582(8): p. 1171-1177.
[20] Bader, G.D., M.P. Cary, and C. Sander, Pathguide: a pathway resource list. Nucle-ic acids research, 2006. 34(suppl 1): p. D504-D506.
[21] Orchard, S. and H. Hermjakob, The HUPO proteomics standards initiative—easing communication and minimizing data loss in a changing world. Briefings in bioinformatics, 2008. 9(2): p. 166-173.
[22] Hermjakob, H., et al., The HUPO PSI's molecular interaction format—a commu-nity standard for the representation of protein interaction data. Nature biotechnolo-gy, 2004. 22(2): p. 177-183.
[23] Kerrien, S., et al., Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. BMC biology, 2007. 5(1): p. 44.
[24] Orchard, S., et al., The minimum information required for reporting a molecular interaction experiment (MIMIx). Nature biotechnology, 2007. 25(8): p. 894-898.
[25] Consortium, U., Ongoing and future developments at the Universal Protein Re-source. Nucleic acids research, 2011. 39(suppl 1): p. D214-D219.
[26] del-Toro, N., et al., A new reference implementation of the PSICQUIC web service. Nucleic acids research, 2013. 41(W1): p. W601-W606.
[27] Orchard, S., et al., Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nature methods, 2012. 9(4): p. 345-350.
[28] Xenarios, I., et al., DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research, 2002. 30(1): p. 303-305.
[29] Kerrien, S., et al., The IntAct molecular interaction database in 2012. Nucleic acids research, 2011: p. gkr1088.
[30] Ceol, A., et al., MINT, the molecular interaction database: 2009 update. Nucleic acids research, 2009: p. gkp983.
[31] Chautard, E., et al., MatrixDB, the extracellular matrix interaction database. Nucle-ic acids research, 2010: p. gkq830.
[32] Goll, J., et al., MPIDB: the microbial protein interaction database. Bioinformatics, 2008. 24(15): p. 1743-1744.
[33] Brown, K.R. and I. Jurisica, Unequal evolutionary conservation of human protein interactions in interologous networks. Genome biology, 2007. 8(5): p. R95.
[34] Lynn, D.J., et al., InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Molecular systems biology, 2008. 4(1).

[35] Stark, C., et al., The BioGRID interaction database: 2011 update. Nucleic acids research, 2011. 39(suppl 1): p. D698-D704.
[36] Stumpf, M.P., et al., Estimating the size of the human interactome. Proceedings of the National Academy of Sciences, 2008. 105(19): p. 6959-6964.
[37] Venkatesan, K., et al., An empirical framework for binary interactome mapping. Nature methods, 2008. 6(1): p. 83-90.
[38] Ideker, T. and N.J. Krogan, Differential network biology. Molecular systems biolo-gy, 2012. 8(1).
[39] De Las Rivas, J. and C. Fontanillo, Protein–protein interactions essentials: key con-cepts to building and analyzing interactome networks. PLoS computational biolo-gy, 2010. 6(6): p. e1000807.
[40] Prasad, T.K., et al., Human protein reference database—2009 update. Nucleic acids research, 2009. 37(suppl 1): p. D767-D772.
[41] Bader, G.D., D. Betel, and C.W. Hogue, BIND: the biomolecular interaction net-work database. Nucleic acids research, 2003. 31(1): p. 248-250.
[42] Klingström, T. and D. Plewczynski, Protein–protein interaction and pathway databases, a graphical review. Briefings in bioinformatics, 2011. 12(6): p. 702-713.
[43] Prieto, C. and J. De Las Rivas, APID: agile protein interaction dataanalyzer. Nucle-ic acids research, 2006. 34(suppl 2): p. W298-W302.
[44] Razick, S., G. Magklaras, and I.M. Donaldson, iRefIndex: a consolidated protein interaction database with provenance. BMC bioinformatics, 2008. 9(1): p. 405.
[45] Güldener, U., et al., MPact: the MIPS protein interaction resource on yeast. Nucleic acids research, 2006. 34(suppl 1): p. D436-D441.
[46] Pagel, P., et al., The MIPS mammalian protein–protein interaction database. Bioin-formatics, 2005. 21(6): p. 832-834.
[47] Brown, K.R. and I. Jurisica, Online predicted human interaction database. Bioin-formatics, 2005. 21(9): p. 2076-2082.
[48] Rual, J.-F., et al., Towards a proteome-scale map of the human protein–protein interaction network. Nature, 2005. 437(7062): p. 1173-1178.
[49] Ramani, A.K., et al., Consolidating the set of known human protein-protein interac-tions in preparation for large-scale mapping of the human interactome. Genome biology, 2005. 6(5): p. R40.
[50] Persico, M., et al., HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. BMC bioinfor-matics, 2005. 6(Suppl 4): p. S21.
[51] Lehner, B. and A.G. Fraser, A first-draft human protein-interaction map. Genome biology, 2004. 5(9): p. R63.
[52] Kanehisa, M., The KEGG database. Silico Simulation of Biological Processes, 2002. 247: p. 91-103.
[53] Joshi-Tope, G., et al., Reactome: a knowledgebase of biological pathways. Nucleic acids research, 2005. 33(suppl 1): p. D428-D432.
[54] Parrish, J.R., et al., A proteome-wide protein interaction map for Campylobacter jejuni. Genome biology, 2007. 8(7): p. R130.
[55] Turner, B., et al., iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database, 2010. 2010: p. baq023.
[56] Mewes, H.W., et al., MIPS: curated databases and comprehensive secondary data resources in 2010. Nucleic acids research, 2011. 39(suppl 1): p. D220-D224.
[57] Hu, Z., et al., VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. Nucleic acids research, 2009: p. gkp406.
[58] Wu, J., et al., Integrated network analysis platform for protein-protein interactions. Nature methods, 2008. 6(1): p. 75-77.
[59] Kamburov, A., et al., The ConsensusPathDB interaction database: 2013 update. Nucleic acids research, 2013. 41(D1): p. D793-D800.
[60] Cerami, E.G., et al., Pathway Commons, a web resource for biological pathway data. Nucleic acids research, 2011. 39(suppl 1): p. D685-D690.
[61] Schaefer, C.F., et al., PID: the pathway interaction database. Nucleic acids research, 2009. 37(suppl 1): p. D674-D679.
[62] Karp, P.D., et al., The ecocyc database. Nucleic acids research, 2002. 30(1): p. 56-58.

[63] Harrington, E.D., L.J. Jensen, and P. Bork, Predicting biological networks from genomic data. FEBS letters, 2008. 582(8): p. 1251-1258.
[64] Skrabanek, L., et al., Computational prediction of protein–protein interactions. Molecular biotechnology, 2008. 38(1): p. 1-17.
[65] Scott, M.S. and G.J. Barton, Probabilistic prediction and ranking of human protein-protein interactions. BMC bioinformatics, 2007. 8(1): p. 239.
[66] Kuhn, M., et al., STITCH: interaction networks of chemicals and proteins. Nucleic acids research, 2008. 36(suppl 1): p. D684-D688.
[67] Turinsky, A.L., et al., Literature curation of protein interactions: measuring agree-ment across major public databases. Database, 2010. 2010: p. baq026.
[68] Hakes, L., et al., Protein-protein interaction networks and biology—what's the connection? Nature biotechnology, 2008. 26(1): p. 69-72.
[69] Aranda, B., et al., PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nature methods, 2011. 8(7): p. 528-529.
[70] Orchard, S., et al., Submit your interaction data the IMEx way. Proteomics, 2007. 7(S1): p. 28-34.
[71] Pastrello, C., et al., Integration, visualization and analysis of human interactome. Biochemical and biophysical research communications, 2014. 445(4): p. 757-773.
[72] Lage, K., Protein-protein interactions and genetic diseases: The Interactome. Bio-chimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2014.
[73] Lage, K., et al., A human phenome-interactome network of protein complexes implicated in genetic disorders. Nature biotechnology, 2007. 25(3): p. 309-316.
[74] Calderone, A., L. Castagnoli, and G. Cesareni, mentha: a resource for browsing integrated protein-interaction networks. Nature methods, 2013. 10(8): p. 690-691.
[75] Niu, Y., D. Otasek, and I. Jurisica, Evaluation of linguistic features useful in extrac-tion of interactions from PubMed; application to annotating known, high-throughput and predicted interactions in I2D. Bioinformatics, 2010. 26(1): p. 111-119.
[76] Zhu, X., M. Gerstein, and M. Snyder, Getting connected: analysis and principles of biological networks. Genes & development, 2007. 21(9): p. 1010-1024.
[77] Barabasi, A.-L. and Z.N. Oltvai, Network biology: understanding the cell's func-tional organization. Nature Reviews Genetics, 2004. 5(2): p. 101-113.