International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)
Volume 2, Issue 3, March 2015, PP 53-58
ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online)
www.arcjournals.org

A Novel Survey on Protein Interaction Networks

M. Mary Sujatha

Research Scholar, Dept of Computer Science,
Krishna University, Machilipatnam, Andhra Pradesh

Abstract: *Bio informatics is a biological study that uses computer programming to process biological data. Bio informatics makes use of data mining techniques to study protein complexes, metabolic pathways and gene networks. At present Protein Protein interaction network (PPI) is one of the key topic for the development and progress of modern system biology. The objective behind this presentation explores bioinformatics basis, computational usage in bioinformatics, interaction patterns of proteins which include prediction and analysis of PPI networks. It also covers existing methods, tools, databases utilized in building protein protein interaction networks*

Keywords: *Bioinformatics, protein, in vivo, in vitro, in silico, PPI, gene fusion, phylogenic approach*

1. INTRODUCTION

Bioinformatics [1] is the application of computer and information technology to the management of biological information. The classic data of bioinformatics include DNA sequences of genes or full genomes. Amino acid sequences of proteins and three-dimensional structures of proteins nucleic acids and protein–nucleic acid complexes. Computers are used to gather, store, analyze and integrate biological and genetic information which cannot be well explained through conventional methods. They are in turn applied to gene-based drug discovery and development. It is considered as an interdisciplinary research area between computer science and biological science. By applying informatics techniques to understand and organize the information associated with molecular biology. bio informatics plays a key role in development of algorithms which are useful in measuring sequence similarity. A fundamental goal of computational biology is the prediction of protein structure from an amino acid sequence.

1.1. What is a Protein

All living organisms consist of living cells and share basic cellular mechanisms. A cell is a smallest structural unit of an organism that is capable of independent functioning. Each cell carries DNA to hold information on how a cell works.DNA makes RNA through transcription ,which transfers short piece of information to different parts of cell .RNA in turn makes protein through translation, A protein is large molecule consisting of amino acids which our bodies and the cells in our bodies need to function properly. Our body structures, functions, the regulation of the body's cells, tissues and organs cannot exist without proteins. Protein accounts for 20% of total body weight. The human body is made up of approximately 100 trillion cells - each one has a specific function. Each cell has thousands of different proteins, which together make the cell do its job. Fig-1illustrates existence of different protein structures such as primary secondary and tertiary structures. In primary structure the linear arrangement of amino acids in a protein and the location of covalent linkages
such as disulfide bonds between amino acids. In secondary structure areas of folding or coiling within a protein. In tertiary structure final three-dimensional structure of a protein, which results from a large number of non-covalent interactions between amino acids.

Protein–protein interactions can be defined as the purposeful physical contacts established between two or more proteins as a result of biochemical events or electrostatic forces. A protein is estimated to interact on average of 10 to 20 proteins. According to gene neighborhood approach [2] if there is neighborhood between genes we can infer that their protein products are likely to associate with one another. According to gene fusion approach [2] it infers protein interactions from protein sequences in different genomes. There is a chance of having homolog for one gene protein to other gene that make a protein chain. According to phylogenic approach [2] proteins interact with one another to perform their function simultaneously.

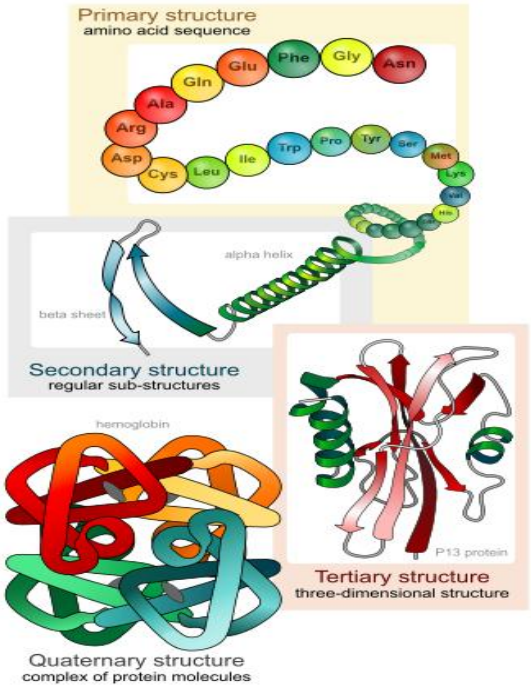

Fig1

2. PROTEIN-PROTEIN INTERACTION NETWORK

Proteins are large, complex molecules that play important role in the body. Proteins do most of the work in cells and are necessary for the function, structure and the regulation of the body. PPI network analysis provides a valuable framework for better interpretation of the functional organization of the proteome. A Protein–Protein Interaction (PPI) is a bio molecules relationship between proteins that plays a foremost role in biological activities. PPI analysis is an area combining bioinformatics and structural biology to rectify interaction between groups of protein. PPI patterns are useful in discovery and development of novel drugs.

Protein networks to diseases reveals:

1. Identifying new disease genes
2. Study of their network properties
3. Identifying disease related sub networks
4. Network based disease classification.

PPI are formally identified as either stable or transient and both types of interaction can be either strong or weak. Multi genetic diseases are mostly associated with the protein complexes due to this reason PPI networks must be evaluated. Two proteins interact with each other if they share a physical binding interface called binary interaction and it can be detected with yeast two hybrid (Y2H)[15] approach. Y2H enabled systematic screening of pair wise PPI. Two proteins interact when they are subunits of same complex called co complex membership. It can be detected with pull down experiments.

2.1. Reconstruction of PPI Network

Reconstruction of protein protein interaction network [14] through multiple sources of evidence. such as interactome maps produced by different labs, other binary maps such as genetic interaction maps, or other genomic features suggestive of protein– protein interaction, taking into account data quality, redundancy and similarity among different sources of evidence leads us to further scope of investigation . Computational approaches are useful in detecting protein complexes and how different proteins are organized into higher level substructure to perform various cellular functions.

3. PPI CLASSIFICATION

A protein never acts alone but interacts with other proteins to form cellular networks. PPIs are classified in several ways [17], Based on their interaction PPIs may be homo or heterogeneous as judged by their stability, they may be obligate or non obligate, as measured by their persistence, they may be transient or permanent. Protein-protein interaction detection methods are categorically classified into three types, namely, in vitro, in vivo, and in silico . According to *in vitro* [2] techniques, a given procedure is performed in a controlled environment outside a living organism. According to *in vivo*[2] techniques, a given procedure is performed on the whole living organism itself. The *in vivo* methods in PPI detection are yeast two-hybrid (Y2H) .According to *in silico*[3] techniques are performed on a computer (or) via computer simulation. Through Y2H and other *in vitro* and *in vivo* approaches are useful in development of tools for detecting protein protein interactions ,the data generated by these approaches are not reliable because of non availability of PPIs. In order to have better understanding on these PPIs. Such that to predict full range of interactions the computational methods of *in silico* are useful. *in silico* speeds the rate of discovery and reduces the need of expensive lab work .In the following section table1 lists different techniques used in said approaches.

3.1. Computational Methods Useful In PPI Analysis

Computationally, protein-protein interaction networks can be conveniently modeled as undirected graphs, where the nodes are proteins and edges represent physical binding interactions. Initially, this graph is missing many edges and contains many incorrect edges. In recent day's data mining techniques such as classification method, Association method, Bayesian network modeling etc., are playing a vital role in computational methods. Classification method uses data to analyze a program to separate positive examples of interacting protein pairs with negative examples of non-interacting pairs. Famous classifiers used are Random Forest Decision (RFD) and Support Vector Machine (SVM). Association

methods emphasis on characteristic sequences or motifs that can help to discriminate interacting and non-interacting pairs. Bayesian Network Method extracts data from different sources to assess a protein interaction is true positive outcome. Computational methods have been successfully applied

1. To compare gene sequences for similarity
2. To identify gene structure.
3. To predict protein structural elements.
4. To navigate genome maps.

These approaches may be categorized on the basis of the types of data they considered when making predictions.

Table1. *List of PPI detection approaches*

Approach	***Technique***	***Summary***
in vitro	Tandem affinity purification-mass spectroscopy (TAP-MS)	TAP-MS is based on the double tagging of the protein of interest on its chromosomal locus, followed by a two-step purification process and mass spectroscopic analysis
	Affinity chromatography	Affinity chromatography is highly responsive, to detect weakest interactions in proteins. It also tests all the sample proteins equally for interaction
	Coimmunoprecipitation	Coimmunoprecipitation confirms interactions using a whole cell extract where proteins are present in their native form in a complex mixture of cellular components
	Protein microarrays (H)	Micro array-based analysis allows simultaneous analysis of thousands of parameters within a single experiment
	Protein-fragment complementation	Protein-fragment complementation assays (PCAs) can be used to detect PPI between proteins of any molecular weight and expressed at their endogenous levels
	Phage display (H)	Phage-display approach is originated in the incorporation of the protein and genetic components into a single phage particle
	X-ray crystallography X	X-ray crystallography provides visualization of protein structures at the atomic level and enhances the understanding of protein interaction and function
	NMR spectroscopy	NMR spectroscopy can even detect weak protein-protein interactions
in vivo	Yeast 2 hybrid (Y2H)	Yeast two-hybrid is typically carried out by screening a protein of interest against a random library of potential protein partners
	Synthetic lethality	It is based on functional interactions rather than physical interaction
	Ortholog - based sequence approach	It is based on the homologous nature of the query protein in the annotated protein databases. It uses pair wise local sequence algorithm
	Domain-pairs-based sequence approach	Domain-pairs-based approach predicts protein interactions based on domain-domain interactions
	Structure-based approaches	Structure-based approaches predict

in silico		PPI if two proteins have a similar structure (Fig-1)
	Gene neighborhood	If the gene neighborhood is conserved across multiple genomes, then there is a potential possibility of the functional linkage among the proteins encoded by the related genes
	Gene fusion	Gene fusion, also called as Rosetta stone method, is based on the concept that some of the single domain containing proteins in one organism can fuse to form a multi domain protein in other organisms
	in silico 2 hybrid (I2H)	The I2H method is based on the assumption that interacting proteins should undergo co evolution in order to keep the protein function reliable
	Phylogenetic tree	The phylogenetic tree method predicts the protein-protein interaction based on the evolution history of the protein
	Phylogenetic profile	The phylogenetic profile predicts the interaction between two proteins if they share the same phylogenetic profile
	Gene expression	The gene expression predicts interaction based on proteins from the genes belonging to the common expression-profiling clusters are more likely to interact with each other than proteins from the genes belonging to different clusters

Table2. *The list of web servers*

Web server	**Function**
Struct2Net	The Struct2Net server makes structure-based computational predictions of PPIs
Coev2Net	Coev2Net is a general frame work to predict, assess, and boost confidence in individual interactions inferred from a high-throughput experiment
PRISM PROTOCOL	PRISM PROTOCOL is a collection of programs used to predict PPI using protein interfaces
InterPreTS	Inter PreTS uses tertiary structure to predict interactions
PrePPI	PrePPI predicts protein interactions using both structural and nonstructural information
iWARP	iWARP is a threading-based method to predict protein interaction from protein sequences
PoiNet	PoiNet provides PPI filtering and network topology from different Databases
PreSPI	PreSPI predicts protein interactions using a combination of domains
PIPE2	PIPE2 queries the protein interactions between two proteins based on specificity and sensitivity
HomoMINT	HomoMINT predicts interaction in human based on ortholog information in model organisms
SPPS	SPPS searches protein partners of a source protein in other species
OrthoMCL-DB	OrthoMCL-DB is a graph-clustering algorithm identifies homologous proteins based on sequence similarity
P-POD	P-POD find and visualize orthologs to a query sequence in the eukaryotes
COG	COG shows phylogenetic classification of proteins encoded in genomes
BLASTO	BLASTO performs BLAST based on ortholog group data
PHOG	PHOG identifies Orthologs based on pre computed phylogenetic trees
G-NEST	G-NEST is a gene neighborhood scoring tool to identify co-conserved, co expressed genes
InPrePPI	InPrePPI predicts protein interactions in prokaryotes based on genomic context
STRING	STRING[13] database includes protein interactions containing both physical and functional associations
MirrorTree	MirrorTree allows graphical and interactive study of the coevolution of two protein families and asseses their interactions in taxonomic context

3.2. PPI 's Experimental Databases

There is a possibility of hundreds and thousands of interactions in PPI, which need to be collected and stored in specialized biological databases that are continuously updated in order to provide complete intercoms. The first databases were the Database of Interacting Proteins (DIP)[16]. Availability of general databases has been increasing. Databases can be divided into primary databases, meta-databases, and prediction databases. Primary databases are called mother databases in turn divided into nucleic acid database and protein sequence database. The nucleic acid sequence databases consists of complete annotation of all the nucleic acid sequences (DNA and RNA) like information of organism from regions, date on which it is sequenced etc., A protein sequence database consists of information of all the proteins that have been translated from the RNA sequences and the proteins sequenced by methods like N-terminal sequencing. The derived databases which are obtained by making use of the sequence information available in the primary databases are called secondary databases. The major structure databases consist of the structural data of the proteins or DNA whose structure has been determined by either X-ray crystallography or NMR (Nuclear Magnetic Resonance).

Database Name	Description
OPHID	The Online Predicted Human Interaction Database (OPHID) is a web-based database of predicted interactions between human proteins.
MINT	It aims at storing, in a structured format, information about molecular interactions (MIs)
MPPI	Mammalian PPI is a new resource of protein interaction data in mammals.
BIND	Bimolecular interaction network database[12] permits an elaborate description of PPI experimentally derived data
BioGRID	The biological general repository for interaction dataset contains protein and genetic interactions among thirteen different species[6]
HPID	The Human Protein Interaction Database (HPID)[7] provides human protein interaction information pre computed from existing Structural and experimental data
iHOP	The information Hyperlinked over Proteins (iHOP) database can be searched to identify previously reported interactions
IntAct	IntAct [8] provides an open source database and toolkit for the storage, presentation, and analysis of protein interactions.

3.3. PPI Network analyzer tools

There are many tools available which are found to be useful to extract PPI networks, pathways from database, to visualize PPI networks, to compare two PPI networks and to identify graph theoretic properties of PPI networks. In recent days databases are integrated with web servers called **APID** [9](Agile Protein Interaction Data Analyzer) which is an interactive bioinformatics web tool developed to allow exploration and analysis of currently known information about protein-protein interactions integrated and unified in a common and comparative platform. At present it include data from six main source databases such as (BIND, BioGRID, DIP, HPRD, IntAct, MINT). Some other tools are listed as follows:

Path BLAST: It is the network analyzer and search tool for comparing protein-protein interaction network among different species to identify pathways of the protein.

Cytoscape It is an open source tools for the analysis and integration of PPI networks

BiogridPlugin2 It allows importing filtering and analyzing PPIs networks from BioGrid Database.

PPI FINDER It is used to mine human PPIs from abstracts based on their occurrences and interaction patterns in human PPI databases.

4. CONCLUSION

This paper presented various methods, servers and tools available in analyzing PPI networks. The area where much research work taken place in discovery of drugs related to Alzheimer's diseases. The scientists analyzed the biology behind Alzheimer's disease using a global approach. Recently identified the largest network of protein interactions related to Alzheimer's disease [10], which is the sixth leading cause of death in old age people. They reviewed proteins with 1412 interactions predicted among 969 proteins. By the importance of PPI study in drug discovery, while available methods and resources are not able to predict complete interactions much research work is going on to invent new pathways. This survey paper helps the researchers in new inventions on PPI that yet to be explored.

REFERENCES

[1]. V. Srinivasa Rao, K. Srinivas, G. N. Sujini, and G. N. Sunand Kumar review article on "Protein-Protein Interaction Detection: Methods and Analysis" Hindawi Publishing Corporation International Journal of Proteomics Volume 2014, Article ID 147648

[2]. Manali R. Modi, Merry K.P "A Survey on Protein-Protein Interaction Network in

Bioinformatics" IJCA 0975-8887, NCWBCB – 2014.

[3]. Kevin R. Brown and Igor Jurisica, "Online Predicted Human Interaction Database" 2076-2082.doi: 10.1093/bioinformatics/bti273

[4]. Andrew Chatr-aryamontri, Arnaud Ceol, Nucl. Acids "MINT: the Molecular INTeraction database" Res. (2007) 35 (suppl 1): D572-D574.doi: 10.1093/nar/gkl950

[5]. Philipp agel, Stefan Kovac, Matthias esterheld, Barbara Brauner ," The MIPS mammalian protein–protein interaction database" Bioinformatics (2005) 21 (6): 832/834.doi: 10. 1093/bioinformatics/bti115

[6]. C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: a general repository for interaction datasets," *Nucleic Acids Research*, vol. 34, pp. D535–539, 2006.

[7]. K.Han,B.Park,H. Kim, J.Hong, and J.Park, "HPID: thehumanprotein interaction," *Bioinformatics*, vol. 20, no. 15, pp. 2466–2470, 2004.

[8]. H. Hermjakob, L. Montecchi-Palazzi, C. Lewington et al.,"IntAct: an open source molecular interaction database," *Nucleic Acids Research*, vol. 32, pp. D452–D455, 2004.

[9]. C. Prieto and J. de Las Rivas, "APID: agile protein interaction DataAnalyzer," *Nucleic Acids Research*, vol. 34, pp. W298– W302, 2006.

[10]. V Srinivasa Rao, K Srinivas, GN Sunand Kumar & GN Sujin," Protein interaction network for Alzheimer's disease using computational approach", ISSN 0973-2063 (online) 0973-8894 (print)

[11]. S. Pitre, M. Alamgir, J. R. Green, M. Dumontier, F. Dehne,and A. Golshani, "Computational methods for predicting protein-protein interactions," *Advances in Biochemical Engineering / Biotechnology* , vol. 110, pp. 247–267, 2008.

[12]. G. D. Bader, I. Donaldson, C.Wolting, B. F. F. Ouellette, T. Pawson, and C.W. V. Hogue, "BIND—the biomolecular interaction network database," *NucleicAcids Research*, vol. 29,no. 1, pp.242– 245, 2001.

[13]. C. vonMering, L. J. Jensen, B. Snel et al., "STRING: known and predicted protein-protein associations, integrated and transferred across organisms," *Nucleic Acids Research*, vol. 33, pp.D433–D437, 2005.

[14]. Eric Franzosa, Bolan Linghu, and Yu Xia "Computational Reconstruction of Protein–Protein Interaction Networks: Algorithms and Issues", Methods Mol Biol. 2009; 541:89-100. doi: 10.1007/ 978-1-59745-243-4_5.

[15]. P. Uetz, L. Glot, G. Cagney et al., "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae," *Nature*, vol. 403, no. 6770, pp. 623–627, 2000.

[16]. I. Xenarios, Ł. Salw´ınski, X. J. Duan, P. Higney, S.-M. Kim,and D. Eisenberg, "DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions," *Nucleic Acids Research*, vol. 30, no. 1, pp. 303–305, 2002.

[17]. I. M. A. Nooren and J. M.Thornton, "Diversity of protein-protein interactions," *The EMBO Journal*, vol. 22, no. 14, pp. 3486– 3492, 2003.