

# **A Survey on Protein-Protein Interaction Network in Bioinformatics**

**Manali R. Modi**

Department of Computer Engineering  
MEF Group of Institutions, Rajkot  
Gujarat, India

**Merry K. P**

Department of Information Technology  
MEF Group of Institutions, Rajkot  
Gujarat, India

## **ABSTRACT**

Bioinformatics is the practical use of computer technology to the management of biological information. The major objective of bioinformatics is to explore and to understand the biological process. Proteins are part of life which is ramified in cellular process. Protein-Protein Interaction (PPI) is a branch which links bioinformatics and structural biology to deal with the prediction, analysis and visualization of protein 3D structures. The openness of these structural data through computer visualization provides understanding of biological processes which cannot be well enough explained by conventional methods. PPI methodology has become essential for knowing human diseases at a wide range. This survey paper encompasses computational perspective of PPI, which can be a supplement to the research area.

## **Keywords**

Bioinformatics, Proteins, cellular process, structural biology, PPI.

## **1. INTRODUCTION**

Bioinformatics is the application of computer and information technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development [2]. It is a discipline of quantitative analysis of information related to biological macromolecules with aid of information technology. It is considered as an interdisciplinary research area at the interface between computer science and biological science. It is a method of conceptualizing biology in terms of molecules and applying informatics techniques to understand and organize the information associated with these molecules, on a large scale. This area is also sometimes referred as a management information system for molecular biology and has many ample of practical applications. It integrates many areas of computer science, statistics, mathematics and information technology to process biological data from large size disease related database [1]. Complex and efficient machines are used to read in biological data at a much faster rate.

The goals of the bioinformatics may be stated as: 1) It organizes data in such a way that allows to access existing information and to submit new data as they produced. 2) To develop tools and resources that is useful in analysis of the data. 3) To use develop tools to analyze the data and interpret the outcomes in an effective manner [4]. The field of Bioinformatics is currently getting considerable attention among researchers. 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. Intention of this paper is to give the new researchers a brief introduction to this emerging topic of research.

## **PPI INTRODUCTION**

Rest of the paper is organized as follows. The Section II describes Protein-Protein Interaction Network. The Section III describes importance of the Protein-Protein Interaction Network. The Section IV describes experimental methods to analyze the Protein-Protein Interaction Network. The Section V describes other multitude computation methods of Protein-Protein Interaction Network analysis. The Section VI describes Protein-Protein Interaction experimental Database. The Section VII describes Protein-Protein Interaction Network analyzer tools. The Section VIII describes conclusion of the Protein-Protein Interaction analysis.

## **2. PROTEIN-PROTEIN INTERACTION NETWORK**

Proteins are large, complex molecules that play important role in the body [6]. Proteins do most of the work in cells and are necessary for the function, structure and the regulation of the body [2]. A Protein-Protein Interaction (PPI) is a biomolecules relationship between proteins that plays a foremost role in biological activities [5]. PPI analysis is an area combining bioinformatics and structural biology to rectify interaction between groups of protein. It is field of bioinformatics that is related to physical contacts established between more than two proteins as an outcome of biochemical reactions. PPI patterns may aid in discovering new drug targets, and support the development of novel drugs. [2].

PPI network analysis provides a valuable framework for better interpretation of the functional organization of the proteome. It plays a vital role in necessary biological activities. PPIs are pivotal for all biological processes [4]. Unravelling protein-protein interactions (PPIs) are one of the central goals in proteomics, which will decipher the molecular mechanisms underlying the biological functions and thereby help to understand human diseases on a system-wide level. [7]. Composing PPI network gives much new insight into functions of a protein [5]. A contradiction in PPI is to explain how unbound proteins are recognized among a large number of cells [5]. PPI are formally identified as either stable or transient and both types of interaction can be either strong or weak. Stable interaction of protein is associated with the proteins that belong to multi-subunit complexes. A protein may interact shortly and in a reversible manner with other proteins in only certain cellular contexts as it happens with most of the proteins involved in biochemical cascades are called transient interactions [6]. Multigenic diseases are mostly associated with the protein complexes due to this reason PPI networks must be evaluated [1].

PPI analysis has become essential for knowing human diseases at a wide range. The field of proteomics gets extended with ability of higher output methods to investigate and analyze PPI network. For that higher dimension, biological databases are used.

In PPI network study, first, analyze the existing problem in detecting methods and functional modules and then analyze the counter measures in the data preprocess and post-process. Second, introduce some special metrics for distance or graph developed in clustering process of proteins. Third, give a classification system of detecting methods and describe some existing detecting methods in each area. Fourth, list database in use and make performance comparisons of several algorithms by different measurements.

### 3. IMPORTANCE OF PROTEIN - PROTEIN INTERACTION

Gene is the basic unit of heredity. A variety of Genomes are present in the body. Proteins are the working molecules of the cell. The human genome consists of 20000–30000 genes for 500000 various proteins of more than 10000 proteins can be produced by the cell at given time PPI are involved in biological activities which are as follows:

**3.1 Signal transduction:** The activity of the cell is carried by extracellular signals. Signals propagation to inside or along the interior of cells depends on PPIs between the different signaling molecules. This process is called signal transduction.

**3.2 Transport across membranes:** A protein may be carrying another protein.

**3.3 Cell metabolism:** In biochemical processes enzymes interact with each other to produce small compounds other macromolecules.

**3.4 Muscle contraction:** Physiology of muscle contraction comprises of several interactions. Various analytical tools like Novus explorer are used for exploring various muscle contraction and gene related diseases.

Primary requirement of PPI networks analysis includes following points in to consideration:

1. Rectify previously-reported data on PPIs from biological databases.
2. Use multiple tools to predict protein modules from protein sequences.
3. Visualize and analyze data on the 3D structure of PPIs networks.
4. Expand ways in which PPIs have been targeted, using bioinformatics approaches, in drug design.
5. Explains the characteristics of PPIs mediated between different kinds of protein modules

### 4. EXPERIMENTAL METHODS TO ANALYZE PPI

PPI have been studied discretely by biophysical and biochemical methods such as co-immunoprecipitation (co-IP) method [7].

#### 4.1 Yeast two-hybrid screening

Yeast two hybrid screening allows the identification of pair wise PPIs, stipulating non-specific tendencies towards sticky interactions. Yeast cells are divided with two plasmids: the bait and the prey. Transcription of reporter genes does not occur unless bait and prey interact with each other and form a

functional transcription factor. The interaction between proteins can be inferred by the presence of the products resultant of the reporter gene expression [4].

#### 4.2 Affinity purification coupled to mass spectrometry

Affinity purification coupled to mass spectrometry widely detects stable interactions and better indicates functional in PPIs. This method initiates by purification of the tagged protein, which is expressed in the cell usually at concentrations, and its interacting proteins. One of the omnipotent advantageous is the tandem affinity purification. PPIs can then be quantitatively as well as qualitatively analyzed by mass spectrometry using various procedures: chemical incorporation, biological or metabolic incorporation (SILAC), and label-free methods [4].

### 5. COMPUTATIONAL METHODS USEFUL IN PPI ANALYSIS

There are multitudes methods (data mining, soft computing and biological) which are useful in PPI for analyzing and evaluating network is present. Few of them are as follows [3]:

#### 5.1 Phylogenetic Profiling

It identifies pairs of protein families having similar patterns of presence or absence among large number of organisms. It is based on the assumptions that potentially interacting proteins must co-evolve. A Phylogenetic profile is generated for every protein under analysis [5].

#### 5.2 Prediction of Co-Evolved Protein Pairs Based On Similar Phylogenetic Trees

This method uses Phylogenetic trees of protein pairs to determine whether interaction between proteins occurs or not [5].

#### 5.3 Rosetta Stone Method

A Rosetta stone protein is a protein comprises of two combined proteins. It is evaluated that proteins that interact with other have homologous in genomes.

#### 5.4 Inference of the Interaction from Homologous Structure

This method is use to manipulate known protein complex structures to predict and structurally model interactions between query protein sequences. The prediction process usually initiates by engage a sequence based method to search for protein complex structures that are homologous to the query sequences.

#### 5.5 Classification Method

A Classification method is one of the most successful methods of data mining technique. It 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).

#### 5.6 Association Method

An Association method is also one of the foremost methods of data mining. It emphasis on characteristic sequences or motifs that can help to discriminate interacting and non-interacting pairs. A classifier is analyzed by looking for sequence-signature pairs where protein comprises of one sequence-signature, and its interacting partner contains another sequence-signature.

## 5.7 Bayesian Network Modeling:

Bayesian Network Method amalgamates data from a huge variety of sources, including some experimental results and prior computational predictions, and utilizes features to assess the likelihood that a particular potential protein interaction is a true positive outcome. These methods are capable because experimental procedures, the yeast two-hybrid experiments, are extremely noisy and produce more false positives, while these computational methods can only provide situational proof that a particular pair of proteins might interact.

## 5.8 Identification of Structural Patterns

This method raise a library of known protein-protein interfaces from the Protein Data Bank, where the interfaces are defined as pairs of polypeptide fragments that are less than a threshold slightly larger than the Van Der Waals radius of the atoms involved.

## 5.9 Domain-Pair Exclusion Analysis

The domain-pair exclusion analysis identifies specific domain interactions that are difficult to detect using Bayesian methods. Bayesian methods are good for detecting nonspecific indiscriminative interactions. The domain-pair exclusion analysis method computes an E-score which measures if two domains interact. High E-scores shows that the two domains are likely to interact, while on another hand low scores indicate that other domains form the protein pair are more likely to be responsible for the interaction. The disadvantage with this method of PPI is that it does not take into account false positives and false negatives in the experimental data.

## 5.10 Supervised Learning Problem

The problem of PPI prediction can be framed as a supervised learning problem. In this method the known protein interactions supervise the estimation of a function that can predict whether an interaction between two proteins exists or not given data about the proteins.

With the help of any of the different method, protein-protein interaction network are analyzed and studied for the different purpose.

## 6. PPI EXPERIMENTAL DATABASE

Large scale PPIs generated hundreds of thousands interactions, which were collected together in specialized biological databases that are continuously updated in order to provide complete interactomes. The first databases were the Database of Interacting Proteins (DIP). The number of general databases has been increasing. Databases can be divided into primary databases, meta-databases, and prediction databases

### 6.1 Primary databases

It collects information about published PPIs proven to exist via small-scale or large-scale experimental procedure. Some of the Examples are: DIP, Biomolecular Interaction Network Database (BIND), sand Biological General Repository for Interaction Datasets (BioGRID), Human Protein Reference Database (HPRD), IntAct Molecular Interaction Database, Molecular Interactions Database (MINT), MIPS Protein Interaction Resource on Yeast (MIPS-MPact), and MIPS Mammalian Protein-Protein Interaction Database (MIPS-MPPI).

### 6.2 Meta-databases

Generally result from the integration of primary databases information, but also collect some of the original data. Some

of the Examples are: Agile Protein Interaction Data Analyzer (APID), The Microbial Protein Interaction Database (MPID8), and Protein Interaction Network Analysis (PINA) platform.

## 6.3 Prediction databases

It includes PPIs that are predicted by various methods. Some of the Examples are: Michigan Molecular Interactions

## 7. PPI NETWORK ANALYZER TOOLS

There are many tools available which are found to be useful for PPIs network some of them are as follows:

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

APID (Agile Protein Interaction DataAnalyzer) is an interactive bioinformatics tool to allow analysis and exploration of highly known information related to PPI integrated. Cytoscape is an open source tools for the analysis and integration of PPI networks.

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

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

## 8. CONCLUSION

In this paper we have done survey of various computational methods, specialized biological databases and different network analyzer tools for PPI network. The world is behind PPI and this survey paper will be guiding the researchers in interesting directions that have yet to be explored.

## 9. ACKNOWLEDGMENTS

Authors are thankful to Nitul Dutta for his advices and critical proofreading of the manuscript.

## 10. REFERENCES

- [1] Antonio Mora, Katerina Michalickova and Ian M Donaldson, "A survey of protein interaction data and multigenic inherited disorders", *BMC Bioinformatics*, vol. 14, pp. 1-7, 2013.
- [2] Fiona Browne, 1 Huiru Zheng,1 HaiyingWang,1 and Francisco Azuaje, "From Experimental Approaches to Computational Techniques: A Review on the Prediction of Protein-Protein Interactions", *Advances in Artificial Intelligence, Hindawi*, vol. 2010, pp. 1-15, 2010.
- [3] Tord Berggård<sup>1</sup>, Sara Linse<sup>1</sup> and Peter James<sup>2</sup>, "Methods for the detection and analysis of protein-protein interactions", *Proteomics*, vol.7 2007, pp. 1-10, 2007.
- [4] Ulrich Stelzl, Uwe Worm, MaciejLalowski, Christian Haening,"A human protein-protein interaction network: A resource for annotating the Proteome", *Cell Press*, Vol.122, issue 6, pp. 957-968, 2005.
- [5] Joan Planas-Iglesias,JaumeBonet,Javier Garcia-Garcia,Manuel A. Marin-Lopez,ElisendaFelliu and BaldoOliva, "Understanding Protein-Protein Interactions Using Local Structural Features", *JMB Article*, pp.1-3, 2013.
- [6] Marco Wiltgen, "Structural Bioinformatics: From the Sequence to Structure and Function", *CurrentBioinformatics*,pp. 1-2, 2009.
- [7] Hung Xuan Ta and Liisa Holm, "Computational approaches for predicting protein interaction networks: The wiring of protein", *The Biochemical Society*,pp.1-3,2011., S. Mullender