A Survey on Protein Protein Interactions (PPI) Methods, Databases, Challenges and Future Directions

Hina Umbrin
Department of Computer Science
COMSATS Institute of Information Technology, Sahiwal
Sahiwal, Pakistan
hinaumbrin88@gmail.com

Saba Latif
Department of Computer Science
COMSATS Institute of Information Technology
Sahiwal, Pakistan
ksaba1595@gmail.com

***Abstract*—Protein Proteins Interactions (PPIs) is a process of interacting protein with proteins in order to produce some organic procedures. To make better understanding of recognizing protein work it is necessary to create high throughput strategies for distinguishing PPI. Protein protein interaction is used for different cells also it defines the 3D structure of proteins. It facilitates different cell capacities resulting from 3D structure of cell that is three dimensional structures of protein protein complexes interactions and it binds affinity information together. The purpose of PPIs is to make interactions between bacterial, viral, and parasitic pathogens of human host's harbors which have great medicinal making potential in order to reduce the causes of dieses that occur due to the PPIs in humans. PPIs are used to discover target specific disease with related interfaces that underlines human interaction network. The structure of ligand binding proteins has to face several challenges including protein sampling of the huge possible orientations for ligand, the protein pocket, sequence space of large data and estimating the binding free energies accurately during the design process. Computational methods are used for successful ligand binding protein design their pros and cons and the potential future directions of the field are discussed. In this paper we have analyzed different methods and techniques of PPIs identification, management, interactions and bindings also to gather different analysis and results based on big databases. In future, we will use Spark's distributed machine Learning library for PPIs prediction, data modeling and machine learning.**

Keywords—PPIs, Machine Learning, Spark, Techniques, Survey, Challenges

I. INTRODUCTION

Proteins are very large, compound units which are assumed to be abundant component in the human body. The formation of the structure and capacity of the cell protein plays an important role in PPIs. There are 20 unique sorts of amino acids that can be merged to make a protein. In every protein's exceptional amino acids chooses each 3-dimensional structure within a specific limit. The development of PPIs in macro molecular structure and enzymatic complexes is important [1]. PPIs control extensive scope of organic procedures, such as cell to cell interactions and metabolic and development control [2].Protein protein interaction (PPIs) are the physical contacts between at least two protein molecules resulting high specificity built up because of biochemical event which is guided by electrostatic powers and the hydrophobic impact. Many physical contacts are molecular relationship between chains that happen in a cell or in a living organism form in a particular bimolecular context. There are different methods for PPIs prediction. The predictions are made by structure based threading approach. Take two protein sequences as one sequence against other sequences from a species. Structure based strings interaction prediction method find all the protein complexes in the Protein Data Bank (PDB) and choice the best potential match. In order to make decisions on PPIs previous results Machine Learning approach is used. Machine Learning is related to artificial intelligence because an intelligent system should be able to adopt change in environment and to store it for further predictions.

Rest of the paper is structured as below. In section II, an introduction and importance of PPIs is presented. The proposed techniques and methods of handling PPIs are addressed in section III. Related work is discussed in section IV. Future work is formulated in section V. Finally, conclusion is given in section VI.

II. RELATED WORK

The different PPIs modeling techniques have been studied by researchers in different platforms including, an auto encoder an artificial neural system that applies an unsupervised learning algorithm which deduces a capacity to develop in order to make conceal structures from unlabeled information [3]. A distributed functional algorithm on Hadoop platform is proposed for PPIs prediction using protein networks [4]. The most precise structure of protein complexes are provided by X-ray crystallography and NMR spectroscopy but these techniques are labor intensive and time consuming[5]. In order to improve the prediction of protein docking sites for a distributed computing architecture, the Map Reduce technique [6] is presented. The majority of the PPIs information was gathered by test strategies, for example, yeast two hybrid (Y2H) screen, tandem affinity purification(TAP) and mass spectrometric protein complex ID (MS-PCI) and other high

978-1-5386-1370-2/18/$31.00 ©2018 IEEE

Throughput procedures to gather PPIs information [7]. A pVLASPD algorithm is purposed for increasing efficiency and effectiveness in order to handle the problem at large scale PPI prediction [8]. Different arrays of distributed computing system (DCSs) have been emerged to handle big data applications. Hadoop and Spark are used to offer better performance when training machine learning (ML) models, a lot of ML systems have been proposed which are based on Parameter Server (PS) framework such as MxNet, MPI-Caffe, TensorFlow, and Petuum. These systems decompose an application in to a set of small tasks and execute them on multiple nodes parallel [9]. Spark is exceptionally adaptable to fit in real environment, tolerating all operations of HDFS [10]. The main purpose of this paper is to utilize Machine Learning techniques to make predictive data models and to highlight their previous working.

A. DataBases

The identified databases in literature proposed by different authors including: Agile Protein Interaction DataAnalyzer(APID)[11], BioGrid [12], ConsensusPathDB (CPDB) [13], IntAct - Molecular Interaction Database [14- 15], Interologous Interaction Database(I2D) [16], Mentha: The Interactome Browser [17], Molecular INTeraction database (MINT) [18], HomoMINT [19], Protein Interaction Network Visualizer (PINV) [20], StringDB - Search Tool for the Retrieval of Interacting Genes/Proteins [21] and Unified Human Interactome (UniHI) [22]. To shortly mention representative tools in the field, 2D standalone applications like graphVizdb [23], Ondex [24], Proviz [25], VizANT [26], GUESS [27], UCINET [28], MAPMAN [29], PATIKA [30], Medusa [31],Osprey [32] as well as 3D visualization tools such as Arena3D [33-34] and BioLayout Express [35] already exist.Several tools are tested including, Cytoscape (v3.5.1) [36], Tulip (v4.10.0) [37], Gephi (v0.9.1) [38], and Pajek (v5.01) [39-40] standalone applications to be the top four candidates for visualization, manipulation, exploration, and analysis of very big networks. Some of these databases and their working are described in detail.

1) Agile Protein Interaction Data Analyzer (APID)

Protein's interaction in graphs is allowed by APID which is represented in a separate Java applet called ApinBrowser. A multi stage tool is prepared with Java applet in which several queries are permitted by APID and results are briefly obtained. Point by point overview of PPIs is done which includes the investigation and data from different interactions. A Pin Browser gives Zoom filter options that limit the detail elements. Add and Import functionality inside the applet makes it possible to visualizing the graph and adds proteins. The UI of queries include a short brief tabular summary of results [41].

2) BioGrid

BioGrid gives a straightforward choice offering a fast look on filter and sorting features which supports multiple platforms. The outcomes indicate essential data in the form of graph that cannot be sent out. Fundamental textual list can be downloaded. A precise organized graph for Cystoscope or alike apparatuses are rejected. The Low to mid-range number of interactions is started using single protein query. Multi protein search is not available in input options. Visual analysis is limited in Bio Grid [42].

3) Consensus Path DB (CPDB)

CPDB uses short computational loading time for intrinsic search in order to combine the required results. For filtrations CPDB is used as a mapping standard which is supportive to PPI analysis tool. CPDB offers many possible outcomes and include many data sources. The graph is great extent and densely packed because of automatic stretching. The visualization of filter function is not assimilated but must be characterized mapping of interactions. There are different criteria for mapping resources, for example combine different results into a specific database. Due to network outlook make it possible to makes use of repositioning and zoon and for highlighting attributes use shading and shape contrasts of nodes and edges. CPDB's graph introductions encourage exploration [43].

4) Molecular Interaction Database (MINT)

A Java installed browser is needed by MINT to execute a query. The search User Interface brings a short review of results including an overview of the different databases utilized. A graphical representation is missing for a fast take for color or shape. Mint demonstrates minimal number of communications for the single protein. Information on related diseases is open showing three connecting proteins out of 93 to be required with obsessive systems. It is prescribed to change to the more forward PPI apparatus upheld by Uniroma Mentha, which offers new portrayal highlights not constrained to Java any more [44].

5) Protein Interaction Network Visualizer (PINV)

The graph representation keeps running in present day browser having JavaScript introduced and enacted.To make a HTML5 application as applied into the tool offers interesting possible results for supporting using the BioJS and D3 framework. Smooth results could be possible at the correct point when input is regularly missing. There are different choices to conveying the diagram both graphically and as literary tables. A fitting enlightening informational index must be select online available sources previously coordinating protein seeks. By picking the human instructive set the single protein input comes to fruition into a higher count of 95 PPIs [45].

6) String DBSearch Tool for the Retrieval of Interacting Proteins/ Genes

String DB's intelligent system watcher is required a present day browser including the Flash module. The query alternative is simple and contains information from a few databases including various organisms. There are possibilities for simple visualization, for example, the occurrence view. String DB gives the other choice to get extra information on disease associations.13 affiliations are found inside the 37 interacting proteins. It bolsters PPI representation and examination [46].

7) iHOP

The iHOP (Information Hyperlinked over Proteins) database is used to search and identify previously reported interactions in PubMed for a protein of interest. It is more efficient than the use ofconventional keyword searches in PubMed [47].

B. Machine Learning

Machine learning algorithms are used to store previous identified patterns of PPI interactions and to predict future patterns of interactions for this purpose stacked Auto Encoder scheme is used. Support vector machine algorithms (SVM) that are trained carefully which picks positive and negative training sets of protein interactions in order to get desired output.

1) Stacked Auto Encoder

Unsupervised learning algorithm is used to create hide structure from unlabeled data which is applied by an auto encoder artificial neural network. An encoding and decoding process takes x as input and make output x^. Different layers of auto encoder are trained layer wise that is covered by SAE. The previous layer will undoubtedly contribute of the dynamic layer other than the yield. The yield is used with delicate max classifier to calibrate all the previous parameters utilizing a back propagation algorithm with classification errors. The structure of a stacked auto encoder model accomplished the highest accuracy 97.9% expectation correctness's for the three external test sets were essentially better. Through a group of amino acids and domain protein interact with one another SAE algorithm of its effective generalization on the protein input sequence limits to learn hidden interactions features [3].

C. Schoring Schemes

Scoring techniques have evolved by combining various features and properties. The methods such as GRAMM- X [48], ATTRACT [49-50], 3D-DOCK [51-52], LZERD are types of scoring scheme.

1) Contact Prediction

Contact prediction techniques discover residue combines that is in the native structure of a protein spatial proximity. Contacts can be used as impediments to oversee abinitio strategies and to redesign the 3D structure of a protein. Assembling is one basic approach in machine figuring out how to consolidate a few wellsprings of information. It uses different models each getting particular parts of the data. A special Meta forecast strategy called EPSILON-CP (combining evolutionary, physicochemical and sequence based information for contact prediction, epsis extend to epsilon) in view of profound neural system that consolidate succession based, formative, and physicochemical information [54].

2) HDOCK SERVER

HDOCK a web server of half and half docking calculation of format based demonstrating and free docking in which deceiving layout can be protected by free docking protocol. Several underpin protein and protein DNA/RNA docking and discover level with arrangement and structure contributions from protein. The HDOCK process is quick and expend 10 to 20 minute for a docking run. The execution of HDOCK grew Better expectations. The HDOCK server is a general edge work for protein and protein DNA/RNA docking that is like half breed docking pipeline work for CAPRI. Compared to show docking servers HDOCK server acknowledges both structure and arrangements as contribution for protein and consequently incorporate the coupling data from the PDB.HDOCK utilizes natural scoring capacity for both protein and protein DNA/RNA docking. HDOCK is a joined bundled of numerous parts including a few outsider programs [55].

III. APPLICATIONS

A. Protein protein interactions in drug discovery

Protein protein interaction is significant to the formation of macromolecular structure and enzymatic complexes. In the recent years PPIs have been evolving as promising targets for rational drugs design due to their high specificity possibly enables the researchers to target specific disease related pathways. There are two types of experimental methods that expose the molecular recognition mechanisms of different types of PPIs methods used for screening large scale and some are used for examine individual PPIs such as high throughput methods as yeast two hybrid system. The methods such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cry electron microscopy are used to individually investigate a specific PPI of interest. These methods are also proficient of determining PPI sites at the atomic level. Such experimental method has some limitation. Because of various physicochemical issues such as transient dynamics, post translational modification (PTM). Needs for in silico approaches to efficiently identify PPIs and PPI sites to expand PPI coverage and filter out the false positives based on confidence scores of interactions between proteins. There is a resilient correlation between network topology and its functioning. A substantial number of in silico prediction methods have been proposed which is used for distinguish PPIs and non PPIs or PPIs sites and non-PPIs sites. Categorize prediction method based on following features sequence, structure, homology, domains, functional similarity, gene co expression and network topology and their potential applications in rational drug design, hotspot prediction and docking. PPIs have attracted attention as drug targets even including single residues the well characterized PPI interactome consist of different structural classes with diverse druggability measure. The pharmaceutical industry remains hesitant to use PPIs as drug discovery. PPIs are difficult to measure enzymatic activity. Rational design of PPI inhibitors depend on protein structures and computational analysis has also spurred progress. The rational design of drugs begins by identifying and characterizes the PPI targets. The molecular level interaction details of the PPI interfaces are critically important for the identification of small molecule modulators. The structural details of the interfaces can be revealing with the availability of in silico structures corresponding to different states. The drug ability and hotspot analysis is used in vitro or in silico biophysical methods that can be targeted by

978-1-5386-1370-2/18/$31.00 ©2018 IEEE

small molecule modulators. The drug ability and hotspot analysis is done using web services. Knowledge of PPI sites can be used to verify PPIs through mutagenesis experiments to find important binding hotspots that contribute to the binding free energy. These experimentally verified PPI sites and hotspot residues can be used to design small molecules to regulate therapeutic and druggable PPIs their effects. Arrange depiction and examination not simply gives a framework level comprehension of medication activity and infection multifaceted nature [56].

B. Proteome scale interactome network maps

Macromolecular interaction network maps are characterized to determine how reported interactions affect the organization, dynamics, and functions of a given biological system. In the era of genomic sequencing, more genetic mutations have been identified than have been functionally characterized. Although over 100000 distorted alleles associated with Mendelian disorders, complex diseases, and cancer have been Catalogued. It is critical then to discriminate between disease causing and natural genomic variations and to determine their respective PPI patterns. A better understanding of disease mechanisms can also be reached through the study of network topology which allows the identification of patterns in interactome networks. One approach known as the disease module hypothesis is based on the observation that disease proteins are not scattered randomly in the interactome, but form topological modules where they tend to interact more with each other than with proteins outside of this neighborhood. These particularly well connected subgraphs of proteins are called disease modules. Complementary to interspecies resource maps, interspecies maps of pathogen host interactomes have been built to study the global landscape of host perturbations by pathogens. A proteome- scale map method used for generating different technique. In which Binary interactions identified using Y2H as the primary screening method are validated by a number of orthogonal assays. This screening method has therefore proven to be a useful tool. In interactome mapping by AP MS, epitope tags are fused to bait proteins. In contrast to Y2H and AP-MS, interactome mapping by co fractionation does not require exogenously introduced ORFs or protein tags. All computational methods rely on different kinds of experimental data. A proteome-scale map serves as a reference allowing for the enumeration of interactions for wild type proteins. Efforts in interactome mapping, with integration of isoforms and protein variants as well as quantitative, spatial, and temporal information, will permit a better understanding [57].

IV. CHALLENGES

There are two major sampling challenges to P-P docking speed and conformational flexibility. Docking techniques must to have the capacity to filter through billions of conceivable setups. Numerous strategies thus utilize FFT based examining as it is swifter than Monte Carlo and geometric fitting based techniques. Project such as FRODOCK and HADDOCK have further improved sampling speed with parallel implementation of their algorithms. In all these methods, the increase in speed does not compromise on accuracy. The problem with conformational flexibility, however, still remains. Methods such as HADDOCK, that take some flexibility into consideration, although at a refinement stage, tend to perform better than methods that treat interactors as rigid bodies. Note that in methods that account for flexibility, just little conformational changes are tested. It is broad conformational variations. For example those observed in some instigated fit interactions remain computationally illusive. To recognize a binding occasion, a quick and precise scoring function should complement an exhaustive testing algorithm. Preferably, the scoring scheme should compute the free energy of binding. Such calculations are difficult to accomplish and none of the present technique consistently make accurate estimates of the binding free energy. Most scoring functions discriminate between binding and non-binding events, while being not able to consistently rank order binding partners by affinity. Another aspect of scoring that requirements improving, in the face of the expanding number of multi segment complexes being modeled, is deciding the relative stoichiometry of binding partners. In other words we need to determine how many concurrent binding partners a particular protein is likely to have. The scoring scheme should be able to decide whether the free energy of the complex would decrease with/without additional interactors. In general, the environment of the interaction could influence the strength of binding. A recent method makes use of explicit water in the docking protocol to account for this environment. The CAPRI experiment serves as an perfect testing platform for new docking technique. It is also a indicator of the progress we have made in protein–protein docking over the years. Expectation challenges are typically connected with interfaces that are made of more than 1 surface patch, or with interfaces that are adaptable. The overall accuracy with which the 3D structures of protein complexes have been modeled has continuously improved over the years.

V. FUTURE DIRECTIONS

- Automatic demonstrating of DNA/RNA structure from arrangements is as yet difficult issue.
- Structural basis for the discovery of a new drug.
- Scoring complexes to determine binding free energy is still a challenging problem.
- Evaluation of features for PPI prediction.
- Comparative Study about Silico technique.
- Researchers which are intriguing in finding new structure of problem area buildups could utilize the HEP model to portray the part of their features.HEP would likewise profit by these new highlights then again
- The need to develop a solid negative data set with wide coverage of protein for PPI prediction to increase the absolute number of PPI samples for training. This idea agrees with the concept of big data which highlights data complexity beside of data volume.

978-1-5386-1370-2/18/$31.00 ©2018 IEEE

Some reflection made for selection of negative samples.

- The accurate prediction of binding affinity of protein protein complexes upon mutation and disease causing mutations at the interface applications will have potential in drug discovery.
- Mechanistic aspects of transcellular BBB trafficking should be an expanding research focus to better inform BBB targeting vector design and optimization.

VI. Conclusion

Different predefined approaches demonstrate how a variety of levels coordinates to make the protein structure and PPIs. These methods permit not only determine how a pathogenic protein interacts with its host on a molecular scale but also how such interactions function in a larger cellular network. Machine learning approaches are used to predict high assurance interactions by combining appropriate sets of positive and negative training sets. In this paper we have surveyed all proposed approaches, methods, challenges, issues and future directions of PPIs and we will solve these challenges by using machine learning techniques in order to predict combinations of PPIs based on learning data.

References

[1] Murakami, Yoichi, "Network analysis and in silico prediction of protein–protein interactions with applications in drug discovery." Current Opinion in Structural Biology, vol. 44, pp. 134-142, 2017.

[2] Rao, V. Srinivasa, "Protein-protein interaction detection: methods and analysis." International journal of proteomics 2014.

[3] Sun, Tanlin, "Sequence-based prediction of protein protein interaction using a deep-learning algorithm." BMC bioinformatics, vol. 18.1, pp.277, 2017.

[4] Akkoyun, Emrah, and Tolga Can, "Parallelization of the functional flow algorithm for prediction of protein function using protein-protein interaction networks." High Performance Computing and Simulation (HPCS), International Conference on. IEEE, 2011.

[5] Szilagyi, Andras, and Yang Zhang. "Template-based structure modeling of protein–protein interactions." Current opinion in structural biology, vol. 24,pp.10-23,2014.

[6] Li, Hui, Jean-Claude Tounkara, and Chunmei Liu. "Prediction of Protein-Protein Docking Sites Based on a Cloud-Computing Pipeline." International Journal of Machine Learning and Computing, vol. 2.6, pp. 798, 2012.

[7] Huang, Yu-An, "Sequence-based prediction of protein-protein interactions using weighted sparse representation model combined with global encoding." BMC bioinformatics, vol. 17.1, pp.184, 2016.

[8] Hu, Lun, "Efficiently predicting large-scale protein-protein interactions using MapReduce." Computational Biology and Chemistry 2017.

[9] Sun, Peng, et al. "Towards Distributed Machine Learning in Shared Clusters: A Dynamically-Partitioned Approach." Smart Computing (SMARTCOMP), IEEE International Conference on. IEEE, 2017.

[10] Karun, A. Kala, and K. Chitharanjan. "A review on hoop—HDFS infrastructure extensions." Information & Communication Technologies (ICT), IEEE Conference on. IEEE, 2013.

[11] Prieto, Carlos, and Javier De Las Rivas. "APID: agile protein interaction DataAnalyzer." Nucleic acids research" vol. 34, pp. W298-W302,2006.

[12] Chatr-Aryamontri, Andrew, "The BioGRID interaction database: 2013 update." Nucleic acids research, vol. 41.D1, pp. D816-D823, 2012.

[13] Kamburov, Atanas, "The ConsensusPathDB interaction database: 2013 update." Nucleic acids research, vol. 41.D1, pp. D793-D800, 2012.

[14] Kerrien, Samuel,"The IntAct molecular interaction database in 2012." Nucleic acids research, vol. 40.D1, pp. D841-D846, 2011.

[15] Orchard, Sandra, "The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases." Nucleic acids research, vol. 42.D1, pp. D358-D363,2013.

[16] Brown, Kevin R., and Igor Jurisica. "Unequal evolutionary conservation of human protein interactions in interologous networks." Genome biology, vol. 8.5, pp.R95, 2007.

[17] Calderone, Alberto, Luisa Castagnoli, and Gianni Cesareni. "Mentha: a resource for browsing integrated protein-interaction networks." Nature methods, vol. 10.8, pp. 690-691,2013.

[18] Licata, Luana, "MINT, the molecular interaction database: 2012 update." Nucleic acids research, vol. 40.D1, pp.D857-D861, 2011.

[19] Persico, Maria, "HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms." BMC bioinformatics,vol. 6.4 ,pp. S21,2005.

[20] Salazar, Gustavo A. "A web-based protein interaction network visualizer." BMC bioinformatics,vol. 15.1,pp. 129 ,2014.

[21] Franceschini, Andrea, "STRING v9. 1: protein-protein interaction networks, with increased coverage and integration." Nucleic acids research,vol. 41.D1,pp.D808-D815 ,2012.

[22] Kalathur, Ravi Kiran Reddy, "UniHI 7: an enhanced database for retrieval and interactive analysis of human molecular interaction networks." Nucleic acids research,vol. 42.D1, pp.D408-D414, 2013.

[23] Bikakis, Nikos, et al. "graphVizdb: A scalable platform for interactive large graph visualization." Data Engineering (ICDE),IEEE 32nd International Conference on. IEEE, 2016.

[24] Köhler, Jacob,"Graph-based analysis and visualization of experimental results with ONDEX." Bioinformatics,vol. 22.11, pp.1383-1390,2006.

[25] Iragne, Florian, et al. "ProViz: protein interaction visualization and exploration." Bioinformatics,vol.21.2,pp.272-274,2004.

[26] Hu, Zhenjun, et al. "VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology." Nucleic acids research,vol.37.suppl_2 ,pp. W115-W121,2009.

[27] Adar, Eytan. "GUESS: a language and interface for graph exploration." *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006.

[28] Borgatti, Stephen P., Martin G. Everett, and Linton C. Freeman. "Ucinet for Windows: Software for social network analysis." (2002).

[29] Thimm, Oliver, "mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes." The Plant Journal, vol. 37.6, pp. 914-939,2004.

[30] Demir, Emek,"PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways." Bioinformatics,vol.18.7,pp. 996-1003, 2002.

[31] Pavlopoulos, Georgios A., "Medusa: A tool for exploring and clustering biological networks." BMC research notes ,vol.4.1,pp. 384, 2011.

[32] Breitkreutz, Bobby-Joe, Chris Stark, and Mike Tyers. "Osprey: a network visualization system." Genome biology,vol. 3.12,pp. preprint0012-1,2002.

[33] Secrier, Maria,"Arena3D: visualizing time-driven phenotypic differences in biological systems." BMC bioinformatics ,vol.13.1 ,pp. 45,2012.

[34] Pavlopoulos, Georgios A.,"Arena3D: visualization of biological networks in 3D." BMC systems biology, vol. 2.1 ,pp.104,2008.

[35] Theocharidis, Athanasios,"Network visualization and analysis of gene expression data using BioLayout Express3D." Nature protocols ,vol. 4.10,pp. 1535-1550,2009.

[36] Shannon, Paul,"Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome research ,vol.13.11,pp. 2498-2504,2003.

978-1-5386-1370-2/18/$31.00 ©2018 IEEE

[37] Auber, David. "Tulip—A huge graph visualization framework." Graph drawing software,pp.105-126 , 2004.

[38] M.Jacomy,T.Venturini,S.Heymann,andM.Bastian,"ForceAtlas2,a continuous graph layout algorithm for handy network visualization designed for the Gephi software,"PLoS ONE,vol. 9, no. 6, Article ID e98679, 2014.

[39] Mrvar, Andrej, and Vladimir Batagelj. "Analysis and visualization of large networks with program package Pajek." Complex Adaptive Systems Modeling,vol. 4.1,pp. 6, 2016.

[40] Batagelj, Vladimir, and Andrej Mrvar. "Pajek-program for large network analysis." Vol.Connections 21.2 ,pp. 47-57,1998.

[41] Jeanquartier, Fleur, Claire Jean-Quartier, and Andreas Holzinger. "Integrated web visualizations for protein-protein interaction databases." BMC bioinformatics, vol.16.1,pp.195,2015 .

[42] Chatr-aryamontri, Andrew, "The BioGRID interaction database: 2017 update.″ Nucleic acids research,vol. 45.D1,pp. D369-D379, 2017.

[43] Kamburov, Atanas, "ConsensusPathDB-a database for integrating human functional interaction networks." Nucleic acids research 37,suppl_1. pp. D623-D628, 2008.

[44] Chatr-Aryamontri,Andrew,"MINT: the Molecular INTeraction database." Nucleic acids research 35.suppl_1, pp. D572-D574, 2006.

[45] Salazar, Gustavo A.,"A web-based protein interaction network visualizer." BMC bioinformatics,vol.15.1, pp.129, 2014.

[46] Szklarczyk, Damian, "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored." Nucleic acids research 39. suppl_1 ,pp.D561-D568, 2010.

[47] .Berggård, Tord, Sara Linse, and Peter James. "Methods for the detection and analysis of protein–protein interactions." Proteomics, vol. 7.16, pp. 2833-2842, 2017.

[48] Tovchigrechko, Andrey, and Ilya A. Vakser. "GRAMM-X public web server for protein–protein docking." Nucleic acids research, vol. 34.suppl_2 ,pp.W310-W314,2017.

[49] May, Andreas, and Martin Zacharias. "Protein–protein docking in CAPRI using ATTRACT to account for global and local flexibility." Proteins: Structure, Function, and Bioinformatics , vol. 69.4,pp.774-780, 2017.

[50] Zacharias, Martin. "Protein–protein docking with a reduced protein model accounting for side-chain flexibility." Protein Science , vol.12.6,pp. 1271-1282,2003

[51] Carter, Phil, "Protein–protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI." Proteins: Structure, Function, and Bioinformatics ,vol.60.2 ,pp. 281-288,2005.

[52] Smith, Graham R., and Michael JE Sternberg. "Evaluation of the 3D-Dock protein docking suite in rounds 1 and 2 of the CAPRI blind trial." Proteins: Structure, Function, and Bioinformatics ,vol.52.1,pp.74-79,2003.

[53] Venkatraman, Vishwesh,"Protein-protein docking using region-based 3D Zernike descriptors." BMC bioinformatics,vol.10.1 ,pp. 407,2009.

[54] Stahl, Kolja, Michael Schneider, and Oliver Brock. "EPSILON-CP: using deep learning to combine information from multiple sources for protein contact prediction." BMC Bioinformatics, vol.18.1, pp.303, 2017

[55] Yan, Yumeng, "HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy." Nucleic Acids Research, 2017.

[56] Murakami, Yoichi, "Network analysis and in silico prediction of protein–protein interactions with applications in drug discovery." Current Opinion in Structural Biology,vol.44 ,pp.134-142, 2017.

[57] Cafarelli, T. M., "Mapping, modeling, and characterization of protein–protein interactions on a proteomic scale." Current Opinion in Structural Biology,vol. 44,pp .201-210, 2017.