

Construction of Protein-Interactions Ontology

Thomas Edison*Health science and medicine Karolinska
Institute, Sweden

Abstract

Protein-protein interaction (PPI) network architectures (maximal complete subnets) are a crucial tool for analysing protein complexes and functional modules. The data deflection from biological studies is complemented by PPI prediction techniques based on cliques. Clique-based prediction techniques, on the other hand, simply consider the network's topology. In a network, false-positive and false-negative interactions frequently obstruct prediction. To address this issue and increase prediction accuracy, we provide an approach that combines the gene ontology (GO) annotations with clique-based method of prediction. We produce two predicted interaction sets that ensure the quality and amount of anticipated protein interactions based on several GO correction algorithms. The PPI network from the Database of Interacting Proteins (DIP) is subjected to the suggested technique, and the majority of the predicted interactions are confirmed by BioGRID, a different biological database. The original protein network is supplemented with the predicted protein interactions, resulting in clique expansion and demonstrating the importance of biological significance.

This research establishes the best technique for building protein-interaction networks that include information derived from protein complexes by merging crystallographic information with protein-interaction data collected through conventional experimental methods. We suggest that a hybrid approach should be taken, in which complexes with five chains or fewer are deconstructed using the matrix model and those with six chains or more are used to derive pairwise interactions using the spoke model. The findings should increase the precision and applicability of studies examining the topology of protein-interaction networks.

Corresponding author:

Thomas Edison

 ThomasEdison45@gmail.comHealth science and medicine Karolinska
Institute, Sweden

Citation: Edison T (2022) Construction of Protein-Interactions Ontology. *Farmacologia y Toxicologia*, Vol. 12 No. 4: 115

Received: 01-Jul-2022, Manuscript No. Ipft-22-115; **Editor assigned:** 04-Jul-2022, PreQC No. Ipft-22-115 (PQ); **Reviewed:** 18-July-2022, QC No.Q-Ipft-22-114
Revised: 23-Jul-2022, Manuscript No. Ipft-22-115 (R); **Published:** 01-Aug-2022, DOI: 10.36648/2174-8365.12.4.115

Introduction

All living things are built on a foundation of protein. Even in a single cell, protein-protein interaction (PPI) is vitally important for all forms of life's development, reproduction, and metabolism. Three techniques can be used to cluster proteins: (I) protein subunit homology, (II) interaction stability, and (III) subunit combination mode. PPI starts the actions of diverse functional or structural proteins in every single cell by joining related proteins. Studying PPI to learn more about protein functions and life activities is an important endeavour because proteins affect various biological processes, even in single cells [1-4].

PPI has been extensively investigated in both experimental and computing contexts. Communal precipitation, Western

blot, and yeast two-hybrid systems are frequently used in PPI experimentation. The two most popular computer approaches for identifying PPI are topology-free approaches and graph-based approaches, which are based on the separation of proteins and specialised clustering methods, respectively. Several computational methods have been developed to identify PPI. Other computer techniques use machine learning to predict PPIs from protein sequences. Bayesian networks were developed to incorporate various genetic features and predict PPIs. Used conjoint triad characteristics derived from sequences to train a support vector machine classifier. In order to extract latent topic features from the conjoint triad features, Pan et al. initially employed a latent Dirichlet allocation model. The learnt topic features were then fed into a random forest classifier to predict

PPIs. Along with the use of various databases and methodologies, researchers have been able to anticipate and investigate PPIs with greater ease and accuracy thanks to the development and invention of computational technologies and the use of updated algorithms [5].

In this study, we used GO keywords and KEGG pathways to analyse an expanded form of PPI (protein-protein function associations). This study aimed to find important GO terms or KEGG pathways that can distinguish between two proteins with and without functional relationships because few PPI studies using computational approaches examined which GO terms were substantially related to the determination of PPIs. With experiment validations reported in Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), a well-known database on collected associations between proteins, as the positive samples, we first extracted protein-protein functional associations. Next, we randomly selected proteins to serve as the negative samples.

Ten datasets were created, each of which had the same positive samples, in order to account for the possibility that the results could be affected by the random selection of negative examples. Using GO keywords and KEGG pathways, each protein-protein functional connection was encoded into a vector. The evaluation of the significance of each characteristic in each dataset was then conducted using mutual information. Some significant features were found from the feature lists, where features were rated in decreasing order of relevance, and their related GO keywords or KEGG pathways were accessible. Finally, to partially understand the distinction between proteins with and without functional linkages, we examined some of the most significant GO keywords and one KEGG pathway [6-9].

Materials and Methods

This section will outline the three-step process for predicting PPIs as well as ways for estimating the expected PPIs. In a PPI network, there are numerous strongly linked regions that frequently interact with functional modules or protein complexes. Multiple strategies can be used to identify strongly related sets. In the beginning, dense zones are discovered using well-known complex detection methods, and the proteins in these regions may interact with one another. We use the adaptive -cores method to break up the dense subnets into smaller sections where the proteins are more tightly connected to increase the likelihood that these proteins would interact [10].

Based on the possibility for loose, nonspecific interactions, systems for ligand/protein binding studies were chosen. Each of these proteins had sizable binding domains that provided enough space for the ligand to attach to the protein in various conformations and at various places. The *E. coli* AcrB multidrug transporter goes through cyclic modifications that first allow a sizable cavity to bind ligands before expelling the contents outside the cell. Because of the cavity's size, all known ligands might theoretically adopt a variety of conformations there. Because of the pump's extreme no selectivity, it is possible that contact with a particular evolutionarily chosen binding site is not necessary for efflux. It also contributes significantly to pharmacology by absorbing various medications and lowering their free content in

the blood. There are at least two sizable pocket domains on the surface of human serum albumin. Many hydrophobic substances, including various steroids, can be bound by steroid transporters, which are also found in the circulation. The pockets of the steroid transporters are larger than what would be needed to bind a straightforward steroid. The known drug interaction sites were explored utilising these model systems.

Four ligands were used to investigate the open binding chamber of AcrB (PDB ID: 3AOD, chain A). The exporter pumps a solvent called toluene. Acridine orange was the first dye used to identify the exporter, and minocycline is an exported antibiotic crystalized with AcrB in the PDB ID: 3AOD structure. Skatole is a poisonous hydrophobic molecule that is common in *E. coli*'s natural environment. Additionally, two sterol-binding proteins were researched (PDB IDs: 1ZHY and 2A1B). Two typical ligand binding pockets are present in the human serum albumin (HSA) protein. It was discovered that Pocket 2 may bind halothane and diazepam (PDB ID: 1E7B, chain A).

Discussion

A dataset of PPIs in *Saccharomyces cerevisiae* that was retrieved from the Database of Interacting Proteins (DIP, version of 2010/6/14) was used to test our methodology. Protein-protein interactions that have been determined through experiment are found in DIP, which is generally regarded as an excellent data source. 26718 interactions are present in this version dataset. We obtain 4997 protein nodes and 23233 protein-protein interaction pairs from the DIP database after removing proteins that interact with themselves.

High-throughput genome-sequencing techniques easily and affordably produce enormous amounts of sequence data. Analyzing individual DNA sequences could result in the creation of customised medications. Proteins are constructed primarily from domains. They perform critical functions in the movement of nutrients, signal transduction, and catalysis in living things. The Pfam domain database is used in this study. Protein domain families with high levels of sequence similarity are present in the database. This method makes the assumption that a specific domain combination corresponds to a protein and that DDIs can mediate protein-protein interactions. It makes sense to investigate protein regulation using DDIs.

The largest mined clique can be up to ten nodes. We decided that a clique's minimum size for PPI prediction should be six in accordance with its maximum size. Clique confidence score cutoff is set at 0.7. 442 PPI predictions are produced using the original DIP dataset's PPI network as a basis. Then, CORE and ALL are generated, respectively, based on the cliques derived from these predictions and various GO rules. 352 predictions are in CORE, and 874 are in ALL. The maximum clique size increases to 16 when these predictions are included in the original dataset, while the number of tiny cliques decreases due to the consolidation of smaller cliques into bigger ones.

Despite the fact that domains are smaller than proteins, it is possible to expand PPIs from the domain level. The enormous amount of protein connections from DDIs and the capability to

look at the accessible PPIs for NoVs result in two major issues. Complex computations are needed for data integration and relationship creation. To get over these challenges, there are cloud computing approaches accessible. Understanding NoVs would benefit from the application of biological knowledge to organism specificity and infectious processes. This might assist scientists in creating treatments for NoVs.

Conclusions

The number, size, and distribution of the nodes in the complexes produced by various protein complex identification techniques vary. Despite the complexes' topological structure, the approach of dissecting them based on -cores can nevertheless locate their dense sections. This research suggests a technique for predicting PPIs that is reliably adaptable to different complexes and the predictions are accurate when combined with the estimations. Predictions made using different complexes identified by various techniques can complement one another and are distinct from those produced using clique methods. In order to fill in the gaps in the protein interaction networks linked to the protein

complexes, the predicted PPIs can be used. The enhanced networks aid in the identification of protein complexes and the investigation of the interactions between proteins in complexes. The number, size, and distribution of the nodes in the complexes produced by various protein complex identification techniques vary. Despite the complexes' topological structure, the approach of dissecting them based on -cores can nevertheless locate their dense sections. This research suggests a technique for predicting PPIs that is reliably adaptable to different complexes and the predictions are accurate when combined with the estimations. Predictions made using different complexes identified by various techniques can complement one another and are distinct from those produced using clique methods.

Acknowledgments

None

Conflict of Interest

None

References

- 1 Bone S, Pethig b (1985) Dielectric studies of protein hydration and hydration-induced flexibility. *J Mol Biol* 181:323-326.
- 2 Takano K, Yamagata Y, Yutani Y K (2003) "Buried water molecules contribute to the conformational stability of a protein. *Protein Eng* 16:5-9.
- 3 Nakamura K, Uhlik MT, Johnson N L, Hahn K M (2006) PB1 domain-dependent signaling complex is required for extracellular signal-regulated kinase 5 activation. *Mol Cell Biol* 26:2065-2079.
- 4 Chakraborty C, Priya Doss G (2014) Evaluating protein-protein interaction (PPI) networks for diseases pathway, target discovery, and drug-design using 'In silico pharmacology. *Curr Protein Pept Sci* 15:561-571.
- 5 L Fu, B Niu, Z Zhu, S Wu (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150-3152.
- 6 Jeong H, Tombor B, Albert R, Oltvai ZN (2000) The large-scale organization of metabolic networks, *Nature* 407:651-654.
- 7 von Mering C, Krause R, Snel (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417:403.
- 8 Theofilatos KA, Dimitrakopoulos CM, Tsakalidis AK, Likothanassis SD (2011) Computational approaches for the prediction of protein-protein interactions: a survey. *Current Bioinformatics* 6:398-414.
- 9 Stark C, Breitkreutz BJ, Chatr-Aryamontri A (2011) The BioGRID interaction database: 2011 update, *Nucl Acids Res*39:D698-D704.
- 10 L. Nanni (2005) Hyper planes for predicting protein-protein interactions. *Neurocomputing* 69:257-263.