International Research Journal of Biochemistry and Bioinformatics Vol. 12(5) pp. 1-3, October, 2022
Available online http://www.interesjournals.org/IRJBB

Mini Review

Protein 3D Structure, Functional Characterization and Sequence Information for Protein-Protein Interaction Prediction

Shahbaz Ali*

Department of Biomedical Sciences, Faculty of Medicine, MAHSA University, Malaysia

*Corresponding Author's E-mail: alishah3434@yahoo.com

Received: 01-Oct-2022, Manuscript No. IRJBB-22-76435; **Editor assigned:** 03-Oct-2022, PreQC No. IRJBB-22-76435 (PQ); **Reviewed:** 17-Oct-2022, QC No. IRJBB-22-76435; **Revised:** 22-Oct-2022, Manuscript No. IRJBB-22-76435 (R); **Published:** 29-Oct-2022, DOI: 10.14303/2250-9941.2022.35

Abstract

The function of a protein limits its evolutionary path through sequence space. Sequence homolog collections document the results of millions of evolutionary experiments in which the protein changes in accordance with these limitations. It is extremely difficult to decipher the evolutionary history contained in these sequences and use it for engineering and prediction applications. Due to the development of low-cost high-throughput genome sequencing, the potential value of resolving this problem has increased (Knop D et al., 2015). It is difficult to separate genuine co-evolution connections from the chaotic collection of apparent correlations. We tackle this problem by inferring residue pair couplings using a maximum entropy model of the protein sequence, restricted by the statistics of the multiple sequence alignment. Unexpectedly, we discover that the strength of these inferred couplings is a very good indicator of the closeness of residues in folded structures. In fact, the highest-scoring residue couplings are remarkably exact and evenly dispersed to characterise the 3D protein structure (Ravi B et al., 2013).

Human proteome sequence variation data may be used to get functional understanding of 3D protein structures. We examined 3D positional conservation in 4,715 proteins and 3,951 homology models utilising genetic variation data from over 140,000 people, employing 860,292 missense and 465,886 synonymous variants. At least one intolerant 3D site is present in 60% of protein structures, as shown by a significant decrease of observed over anticipated missense variation. Data on structural intolerance were connected with shallow mutagenesis data for 1,026 proteins and functional readouts from deep mutational scanning for PPARG, MAPK1/ERK2, UBE2I, SUMO1, PTEN, CALM1, CALM2, and TPK1. Different characteristics for ligand binding pockets and orthosteric and allosteric locations were found by the 3D structural intolerance analysis. A definition of functional 3D locations proteome-wide is supported by extensive data on human genetic diversity (Valverde ME et al., 2015).

INTRODUCTION

The main component of living things is protein. It engages in a variety of biological processes after interacting with other proteins. PPIs, or protein-protein interactions, aid in understanding how proteins operate, the origins and progression of illnesses, and the development of novel medications. The discovery of protein-protein interactions, however, falls far behind the sequences of the proteins that are already known. Researchers put forth a number of computational techniques to shed light on protein interactions in order to close this knowledge gap. These techniques just rely on protein sequence information. With the development of technology, several kinds of information about proteins are now accessible, including data on their three-dimensional (3D) structures. Deep learning techniques are currently being effectively applied in several fields, including bioinformatics (Sánchez C 2010).

Therefore, current research focuses on using many modalities, such as deep learning algorithms and 3D protein structures and sequence data, to predict PPIs. There are various steps to the suggested strategy. Using their 3D

coordinates and three properties, such as their hydropathy index, isoelectric point, and charge of amino acids, we first obtain a number of depictions of proteins. The building components of proteins are amino acids (Oloke JK 2017). These representations of proteins are analysed using a pre-trained ResNet50 model, a subtype of a convolutional neural network, to extract features. Here, another modality of protein sequences is employed, namely autocovariance and conjoint triad, which are two extensively used sequence-based techniques to encode proteins. To obtain sequence-based information in its compact form, a stacked autoencoder is used (Rom O et al., 2016).

In order to predict labels for protein pairings, the information gleaned from several modalities are finally concatenated in pairs and put into the classifier. On the human PPIs dataset and the Saccharomyces cerevisiae PPIs dataset, we conducted experiments and contrasted the outcomes with cutting-edge deep-learning-based classifiers. The proposed strategy yields better outcomes than those produced by the currently used techniques. Numerous tests on various datasets show that our method of mixing and learning characteristics from two separate modalities is effective for PPI prediction (Hampton T 2017).

The space of potential sequences and, consequently, structures compatible with a functioning protein in the context of a replicating organism is continually being sampled by the evolutionary process. Strong selection constraints make it impossible for amino acid changes to take place in certain places, making it possible to identify homologous proteins from a variety of species by comparing their sequences. The harmony between sequence exploration and restrictions in this evolutionary record, as represented in protein family databases like PFAM, is beautiful. Strong limitations on sequence variation are imposed by conservation of function within a protein family, and this typically ensures that all family members have comparable 3D structures (Rom O et al., 2018).

The level of genetic variation in the human population is described in depth by recent large-scale sequencing initiatives of the human genome and exome. The human exome has been found to include about 4.5 million missense (amino acid-changing) variations as of this writing. The connection between genetic variations and illness has received a lot of attention. These data, however, also offer a rare chance to examine protein structure-function correlations in vivo. The genetic variation distribution pattern, in particular, describes the functional constraints on structural and functional changes to a specific protein. The inference of crucial 3D locations may be instructive for drug development and action mechanisms like as selectivity, resistance, and toxicity (Caldow MK et al., 2016).

It has been possible to locate significant areas inside these buildings using a number of different techniques. The deleteriousness of genetic variations in a protein may be assessed using genetics-based scoring metrics; this characteristic is highly correlated with both molecular functioning and pathogenicity. Scores may take interspecies conservation into account to find "constrained components" that may be signs of potential functional elements. The distribution of variations in 3D space has not received as much attention in previous techniques as gene-level characteristics (such as essentiality, burden of variation, etc.) and linear studies of variation in a gene. To evaluate the clustering of somatic variations in protein structures, further techniques have been developed in the field of cancer.

Described Using exome sequence data from The Cancer Genome Atlas (TCGA) from up to 7,215 samples, 23 types of cancer, and over 975,000 somatic mutations, the identification of protein amino acid clustering (iPAC), spatial protein amino acid clustering (SpacePAC), graph protein amino acid clustering (GraphPAC), and quaternary protein amino acid clustering methods analysed 3D position and clustering of mutations. Recently, a comparison of methods for the subgene-resolution identification of cancer drivers was presented. It should be highlighted that scoring techniques used in oncology do not prioritise intolerance to variation in the human proteome as a whole, but rather mutational clustering, which is extremely important in cancer biology (Heresco-Levy U et al., 1999).

Protein serves as the primary building block of all living things. It participates in a variety of biological processes. Hormone control, metabolism, signal transmission, cell transcription, and replication are some of these processes. The majority of these processes involve various protein interactions. Understanding biological processes, developing novel medications, and determining the progression and origins of illnesses are all made possible by the study of protein-protein interactions. Additionally, using gene interaction network analysis and PPI information, it is possible to anticipate therapeutic targets, for instance in the case of harmful microorganisms. PPIs have been discovered using a variety of high-throughput experimental approaches, including tandem affinity purification (TAP), yeast two-hybrid (Y2H), and mass spectrometric protein complex identification.

CONCLUSION

However, these experimental techniques for detecting PPI have several drawbacks, such as the fact that they are expensive and time-consuming, which prevents them from investigating all PPI networks. Additionally, the experimental setting and operational procedures have an impact on these methodologies' results, leading to significant false positives (FP) and false negatives (FN). In order to effectively predict protein-protein interactions, strong computational approaches must be developed in addition to experimental techniques.

Using the most recent technology, researchers have gathered

multi-modal representations of biological data. The order of amino acids is one type of depiction, whereas a 3D picture of a protein's structure is another. These two protein-related modalities each have unique information on proteins that works well together. Deep learning algorithms have recently made it simpler to discover valuable characteristics from several modalities. The availability of multi-modal biological data has already been used by certain researchers in their work. Have employed a multimodal technique to identify protein distant homology

Thus, the first obstacle is to eliminate the impact of confusing elements in order to solve the inverse sequence-to-structure problem. The ability to anticipate the protein fold depends on whether the evolutionary process has revealed enough residue interactions that are uniformly dispersed (spread) throughout the protein sequence and structure. The precision of 3D structure prediction utilising the inferred contacts is the final performance criteria. Previous research coupled a limited number of evolutionarily inferred residue interactions with additional structural sources of information to correctly predict the structures of several smaller proteins. However, there are still three significant unanswered issues with the use of residue-residue couplings deduced from evolution to predict protein fold.

The first is whether it is possible to create an approach that is sufficiently reliable to recognise causal connections that represent evolutionary restrictions. The second is whether the inferred, evolutionary-plausible connections predominantly reflect closeness between residues. Third, without the aid of existing three-dimensional structures, a protein fold be predicted using the inferred residue-residue proximities.

REFERENCES

1. Knop D, Yarden O, Hadar Y (2015). The ligninolytic peroxidases in the genus Pleurotus: divergence in activities, expression, and potential applications. Appl Microbiol Biotechnol. 99: 1025-1038.
2. Ravi B, Renitta ER, Prabha LM, Reya I, Shanti Naidu (2013). Evaluation of antidiabetic potential of oyster mushroom (Pleurotus ostreatus) in alloxan-induced diabetic mice. Immunopharmacol Immunotoxicol. 35: 101-109.
3. Valverde ME, Hernández-Pérez T, Paredes-López O (2015). Edible mushrooms: improving human health and promoting quality life. Int J Microbiol.
4. Sánchez C (2010). Cultivation of Pleurotus ostreatus and other edible mushrooms. Appl Microbiol Biotechnol. 85: 1321-1337.
5. Oloke JK (2017). Oyster mushroom (Pleurotus species); a natural functional food. J Microbiol Biotechnol Food Sci. 7: 254.
6. Rom O, Aviram M (2016). Endogenous or exogenous antioxidants vs. pro-oxidants in macrophage atherogenicity. Curr Opin Lipidol. 27: 204-206.
7. Hampton T (2017). How useful are mouse models for understanding human atherosclerosis? Review Examines the Available Evidence. Circulation 135: 1757-1758.
8. Rom O, Volkova N, Jeries H, Grajeda-Iglesias C, Aviram M (2018). Exogenous (pomegranate juice) or endogenous (Paraoxonase1) antioxidants decrease triacylglycerol accumulation in mouse cardiovascular disease-related tissues. Lipids. 53: 1031-1041.
9. Caldow MK, Ham DJ, Godeassi DP, Chee A, Lynch GS, et al (2016). Glycine supplementation during calorie restriction accelerates fat loss and protects against further muscle loss in obese mice. Clin Nutr. 35: 1118-1126.
10. Heresco-Levy U, Javitt DC, Ermilov M, Mordel C, Silipo G, et al (1999). Efficacy of high-dose glycine in the treatment of enduring negative symptoms of schizophrenia. Arch Gen Psychiatr. 56: 29-36.