Mini Review

Open Access CrossMark

Computational annotation of protein function

Keywords: kinases, elucidation, mycobacterium, Rosetta stone method, Gibbs distribution

Volume 7 Issue 3 - 2018

Khush Bakht Aliaa, Muhammad Hussnain Siddiquea, Ijaz Rasula, Farrukh Azeema, Muhammad Razeen Ahmada, Habibullah Nadeema

Department of Bioinformatics and Biotechnology, Government College University, Pakistan

Correspondence: Habibullah Nadeem, Department of Bioinformatics and Biotechnology, Government College University, Pakistan, Email habibullah@gcuf.edu.pk, habibullah81@gmail.com

Received: December 26, 2017 | **Published:** June 22, 2018

Mini review

In organisms, in order to understand the molecular mechanisms it is necessary to interpret the protein function,[1] because biological cells are controlled by different signaling pathways and metabolic reactions these reactions are carried out by the interaction of different proteins with proteins or other molecules[2] Proteins play important and various roles in organisms as they act as catalysts in different biochemical reactions, transmission of signals and transportation of nutrients. The term 'protein function' is a complex phenomenon. For instance, kinases are related to various functions of cell like cell cycle, and these kinases act as transferase in different chemical reactions. If mutations occur in proteins, they most likely cause diseases. The part proteins play in the cell may be structural or functional; but mostly mutations in functional proteins like proteins in glycolysis pathway, cause disruption of a whole chain of reactions, and these disturbances ultimately lead to diseases. Therefore, the functional proteins of the cell need to be thoroughly studied and maximum elucidation of their function.[3]

The term biological function of a protein is a little bit ambiguous. The exact meaning of this term varies and depends upon the context in which it is used. The term protein function is a broader term [3] and has more than one aspect. There are several computer-based tools which can be used for in silico prediction of protein function just out of protein sequence. While using a protein function prediction tool, it is important to consider which functional aspect is to be predicted.[4]

The knowledge of protein-protein interactions has been accumulated by the experiments of genetic and biochemical experiments.[2] Every year, hundreds of protein sequences and structures are determined, but the experimental verification of protein function having known sequence and structure is a difficult task.[1] In post-genomic era, the challenging problems is determination of function of protein. Today, the attention has been shifted from the study of small complexes and single proteins to the entire proteome, by whole genome sequencing and the possibility to access co-expression patterns of genes.[5] Therefore, it is important to search new and reliable methods to assign functions to proteins.[6] Better and better computational approaches have been paving way for prediction of protein function. These computational approaches allow us to determine the function of whole proteome.[7]

Bioinformatics intervention

The chief goal of bioinformatics is the determination of protein function by genomic sequences. The assignment of function to a gene product comes from biochemical/molecular biology experiments, which extends by matching a recently sequenced protein to already characterized ones.[8] The basis of protein function prediction is the similarity search for the protein with unknown function among the proteins having known function.[7] Two proteins can be considered similar if their sequences and structures match well and they have similar binding sites, similar interaction patterns and certain amino acid motifs.[9] A newly discovered protein is assumed to perform the same function as the similar proteins in a database do. The similarity suggests that similar proteins could have the same ancestor, and will perform similar functions as would have occurred in their common ancestor.[10]

To discover the function of an unannotated protein is a difficult task. For this purpose, a number of methods have been developed. The traditional way to determine the function of protein is sequence or structure homology search (the latter for those proteins whose structure is newly determined) between the unannotated protein and known proteins in different databases. Besides homology, there is another method known as "Rosetta stone method". In this method, two proteins are considered to be similar if their homologs in some other organisms are expressed as a single polypeptide chain. Additionally, the function of a protein could be predicted by its phylogenetic pattern; the genes with similar phylogenetic patterns may have similar functions. In a cell, proteins drive cellular processes through their interaction with other proteins. The pattern of protein-interaction in a pathway is important and may serve in determining the interactions of a protein with unannotated function.

Marcotte et al.,[12] used a method for protein function prediction, based on interactions of proteins that play role in common metabolic pathways—functionally-linked proteins. The idea behind this approach was the two proteins are likely to be functionally-linked if these two are homologous in the same subgroup of organisms. In organisms the phylogenetic profiling of protein reveals the presence or absence of homologs. The comparison of phylogenetic profiles is an attractive tool for the identification of pathway in which a protein participates[8] originally, there were different methods based on frequencies of interaction partners and chi-square statistics used for the assignment of function to unannotated proteins. These methods, however, did not have systematic mathematical model. Therefore, a mathematical model 'Gibbs distribution' was developed to check the probability of an unannotated protein for certain functions of interest.[11]

Starting from Scratch

For complete characterization of a protein, the most common method for protein function prediction is sequence to function method. For sequence to protein function, two methods are available viz. sequence motif and sequence alignment. These methods have been powerful, but as the protein sequence databases are becoming more and more diverse, these methods show certain limitations, and a single analysis is not enough for completely elucidating the function out of an amino acid sequence. Methods including combination of sequence and structure information show success in protein function prediction, but that does not cover all the dimensions; during evolution, proteins may gain and lose function. They may have several functions in the cell and only the sequence to function methods cannot unlock the complexities of protein interaction, which is majorly dependent upon its folding (structure). The alternative method for function prediction is sequence-to-structure-to-function. The purpose of this approach is the determination of protein structure and then identification of key residues that are functionally important in either catalysis or binding. Use of molecular structure for the identification of functional sites is in line with how protein works.[12]

As the starting point of determination of gene function is similarity search by a sequence alignment tool like PSI-BLAST. Apart from sequence similarity, other techniques like structure prediction, clustering of expression data, and combined approaches are in use for inferring the function. Des Jardins et al.,[13] used three machine learning methods for prediction of enzyme class, by utilizing features computed by amino acid sequence from PDB and SwissProt.[13] King et al used data mining by supervised machine learning program C4.5 and inductive logic programming in order to learn rules established on homology, sequence and structure data from E. coli and Mycobacterium tuberculosis genomes[14] Supervised learning algorithms have been used in support vector machines (SVMs) to predict expression of unannotated ORFs of yeast genome.[15] Amanda Clare & Rose D King, et al.,[14] used supervised machine learning to predict functional class of ORF from phenotypic data in Saccharomyces cerevisiae. They modified C4.5 algorithm in order to overcome the problems from machine learning.[16]

Computational biology—key to success

The key to understanding life at molecular level is the correct annotation of protein function. Therefore, in molecular biology and computational biology, the computational annotation of protein function has emerged as a significant problem.[17] The prediction of protein function by computational means is the central undertaking in computational biology. Although the computational methods are effective and resource-saving, their challenges in correct and confident prediction are still debatable[18] Conventional approaches like phylogenetic profiles, sequence similarities, protein-protein interactions and clustering of co-regulated genes[19] have been used to deduce the function of a protein. Recent research for assigning the function of a protein is based on network of physical interactions of proteins.[6] Instead of all this, these rules are not universally valid; some proteins having similar function could also have dissimilar structures, and proteins having similar structures may have different functions. Moreover, a single amino acid mutation may change the function of protein and lead to having closely related structure with another protein but different function. Because of these exceptions, there is no single function prediction system that can predict protein function accurately. The solution of these problems is to integrate protein data from different sources. For instance, if two proteins are said to be similar on more than one scale, then its function prediction will be more reliable.[1] Since, the meaning of biological function of a protein is very contextual, it is important to consider the aspects of biological function while using any computational approach for the prediction of protein function.[18] On organismal and cellular level physiological experiments may provide information about its role in cell, and biochemical function of a newly categorized protein, which tells us about its role in the life of the cell.[18]

Instead of homology assessment alone, new means are required for annotation because of increase in diversity of protein sequences.[3,4] Concerns in the use of computational methods for protein function prediction are vocabulary standardization of function annotation and assessment of function prediction programs. Again, there is no single program that houses all of the algorithms or strategies to tackle a function prediction. In computational molecular biology, the function prediction is a leading problem and it is necessary to know that how good an individual program is performing.[18] With the increase in computational approaches and hardware technologies, better algorithms and programs are expected to provide even better results in structure and function prediction of complex macromolecules like proteins.

Acknowledgements

None.

Conflict of interest

The author declares no conflict of interest.

References

1. Borgwardt KM, Ong CS, Schönauer S, et al. Protein function prediction via graph kernels. *Bioinformatics*. 2005;21(Suppl 1):i47–i56.

2. Marcotte EM, Pellegrini M, Ng HL, et al. Detecting protein function and protein–protein interactions from genome sequences. *Science*. 1999;285(5428):751–753.

3. Rost B, Liu J, Nair R, et al. Automatic prediction of protein function. *Cell Mol Life Sci*. 2003;60(12):2637–2650.

4. Friedberg I. Automated protein function predictionâ€ the genomic challenge. *Brief Bioinform*. 2006;7(3):225–242.

5. Hodgman TC. A historical perspective on gene/protein functional assignment. *Bioinformatics*. 2000;16(1):10–15.

6. Vazquez A, Flammini A, Amos Maritan et al. Global protein function prediction in protein–protein interaction networks. *Nature Biotechnology*. 2003;21:697–700.

7. Whisstock JC, Lesk AM. Prediction of protein functions from protein sequence and structure. *Q Rev Biophys*. 2003;36(3):307–340.

8. Pellegrini M, Marcotte EM, Michael J Thompson et al. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad Sci.* 1999;96(8):4285–4288.

9. Yao H, Kristensen DM, Mihalek I et al. An accurate, sensitive, and scalable method to identify functional sites in protein structures. *J Mol Biol*. 2003;326(1):255–261.

10. Bartlett GJ, Annabel E, Janet M Thornton, et al. Inferring protein function from structure. *Structural Bioinformatics*. 2005;44:387–407.

11. Deng M, Zhang K, Mehta S et al. Prediction of protein function using proteinâ€ protein interaction data. *J Comput Biol*. 2003;10(6):947–960.

12. Skolnick J, Fetrow JS. From genes to protein structure and function: novel applications of computational approaches in the genomic era. *Trends Biotechnol*. 2000;18(1):34–39.

13. Des Jardins M, Karp PD, Krummenacker M, et al. Prediction of enzyme classification from protein sequence without the use of sequence similarity. *Proc Int Conf Intell Syst Mol Biol*. 1997;5:92–9.

14. King RD, Karwath A, Clare A, et al. The utility of different representations of protein sequence for predicting functional class. *Bioinformatics*. 2001;17(5):445–454.

15. Brown MP, Grundy WN, Lin D et al. Knowledge–based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A*. 2000;97(1):262–267.

16. Clare A, King RD. Machine learning of functional class from phenotype data. *Bioinformatics*. 2002;18(1):160–166.

17. Radivojac P, Clark WT, Oron TR, et al. A large–scale evaluation of computational protein function prediction. *Nat methods*. 2013;10(3):221–227.

18. Godzik A, Jambon M, Friedberg I, et al. Computational protein function prediction: are we making progress? *Cell Mol Life Sci*. 2007;64(19):2505–2511.

19. Harrington CA, Rosenow C, Retief J, et al. Monitoring gene expression using DNA microarrays. *Curr Opin Microbiol*. 2000;3(3):285–291.

Citation: Aliaa KB, Siddiquea MH, Rasula I, et al. Computational annotation of protein function. *MOJ Proteomics Bioinform.* 2018;7(3):195–197.
DOI: 10.15406/mojpb.2018.07.00233