Application of Bioinformatics in Predicting Protein-Protein Interaction

Xiaoyang Han

Marquette University, Wisconsin, Usa

rocketmanhan@gmail.com

Keywords: Bioinformatics, Predict, Protein, Interaction

Abstract: In the post genome era, bioinformatics data has been accumulated, and bioinformatics has been applied in many fields. Bioinformatics is a comprehensive subject, integrating the research of biology, life science and statistics. It has high research value, especially in the prediction of protein-protein interaction. Using bioinformatics to study protein-protein interaction can not only achieve high scientific research results, but also play a guiding role in later experiments. This paper mainly explores the application of bioinformatics in predicting protein-protein interaction, hoping to be helpful to medical research.

1. Introduction

With the fast progress of information technology, bioinformatics keeps growing and it is an effective research method in the post genome era. In the post gene era, proteome is very complex and challenging, and there is a great demand for bioinformatics data. At present, the research focus of bioinformatics is the protein-protein interaction. The application of bioinformatics can deepen the understanding of protein-protein interaction and provide a strong basis for revealing the mystery of life activities.

2. Related Concepts of Bioinformatics

Bioinformatics is a comprehensive subject, rising with the start of genome project. It integrates the research of mathematics, computing and information science, including the collection, integration, analysis and processing of biological information, and fully illustrates the biological significance of biological data. Bioinformatics mainly studies protein and other macromolecular databases. By means of mathematics, computer and so on, it processes and arranges a large number of original data, making these bioinformatics have biological significance[1]. Through the comparison and integration of biological information, gene coding and so on can be obtained, which provides a scientific basis for exploring the unknown world.

3. Application of Bioinformatics in Predicting Protein-Protein Interaction

Bioinformatics integrates the research of mathematics, physics and other disciplines, and uses computational methods to study the interaction between proteins, effectively shortening the research cycle, reducing the research cost, and opening up a way for the research of life activities. To predict protein-protein interaction more scientifically, four bioinformatics methods are introduced. Through the introduction of these four methods, we can understand the advantages and disadvantages of each method, so as to provide help for future progress.

3.1 Application of Genome Information Method

There are three prediction methods of genome information: gene fusion, gene adjacency and phylogenetic spectrum. Firstly, gene fusion is a method to predict protein-protein interaction by observing the evolution information of gene structure in the process of gene evolution. The fusion protein can be obtained by linking different genes, and it has a certain complex function. The fusion of genes indicates that there must be some functional association between these genes. In one

Copyright © (2021) Francis Academic Press, UK DOI: 10.25236/mmmce.2021.014

species, two proteins are relatively independent and have their own gene codes, but in other species, they are the same gene codes. It can be concluded that the two proteins are independent of each other, or the same protein is functionally related in different domains. There are basic principles in the composition of fusion protein. To construct the fusion of two genes, the first protein gene can be connected with the second protein gene, and the first stop codon must be deleted, and the second one must have a stop codon. This principle also applies to finding protein-protein interactions. This method is widely used in metabolizing proteins. However, this method also has some shortcomings. It can only predict the functional association between fusion proteins, but can't predict whether these fusion proteins have physical contact. There are few gene fusion events, so the application of this method is limited. Secondly, the gene adjacency method considers that the interacting protein coding genes should be close to each other in the genome. When one of the genes is expressed, the expression of adjacent genes will also appear. This method is very conservative and can be used as an indicator to show the functional relationship of genes. This method is not suitable for eukaryote specific proteins, only for some microorganisms with simple structure and early evolution. Using this method, we can see the association between the two genes, but there is still a lack in predicting the interaction. Although some genes are very close in the genome, they have no relationship. Therefore, this method is more suitable for conservative genome prediction. Finally, in the method of phylogenetic spectrum, evolutionary analysis is the tool to predict protein related functions. This method believes that in the process of gene evolution, the function related genes will appear or lose at the same time. Gene functions can be predicted by observing and analyzing the distribution of genes in different species. The phylogenetic spectrum is similar, but the two gene sequences without homology can also predict that they may have related functions, so the protein functions corresponding to the genes must be related. This method is enlightening and can be used to discover the potential functions of proteins[2].

3.2 Application of Amino Acid Sequence Method

In the post gene era, the research object of life science is mainly protein, and the main task is to study the relationship between the structure and function of protein, in which the composition of amino acids is closely related to the structure of protein. The information in amino acid sequence plays a decisive role in the structure of protein and determines the folding structure of protein. It can be seen that it is very possible to predict the protein structure according to the amino acid sequence information, and then explain the protein-protein interaction effectively. At present, there are many analytical methods to maintain and find the amino acid residues of protein structure. Some scholars have found that amino acid residues in different states play different roles in protein function, and some researchers have found that the active sites of proteins can be found by effectively searching the cracks in protein structure. Many functions of proteins are determined by their structures. It can be said that all information about proteins can be found in the arrangement of amino acids. Through the correlation signal of amino acid sequence, we can effectively identify and label the related proteins, so as to predict the interaction of different functional proteins.

3.3 Application of Evolutionary Information Method

There are two methods of evolutionary information. One is association mutation, and the other is mirror tree. The association mutation means that a whole system is closely integrated. If one part of the system mutates, other parts will also mutate. From a physical point of view, if one of the two proteins in contact has a residue change, it will be compensated from the other system. In order to ensure the structure stability and limit the continuous mutation of genes, the molecular mechanism of related mutation will make sequence adjustment, which is small and not easy to detect. Through association mutation and alignment of multiple sequences, conservative positions can be generated, which has a great effect on increasing probability of predicting and identifying contact points. The results show that correlation variation not only occurs between molecules, but also within molecules. In the application of this method, we need to consider the proteins in the genome. Mirror tree is based on the hypothesis that the interacting proteins are coevolutionary. Through this hypothesis, we can conclude that there is a certain association between the interacting proteins. The association

is either genetic or physical. Mirror tree is a typical method based on evolutionary information. It is based on the hypothesis of coevolution, believing that under the constraints of function, the evolution process of related genes should be the same or similar, with the characteristics of coevolution. The structure of gene phylogenetic tree should be similar. When predicting the related functions of mirror tree genes, we can compare the phylogenetic tree of genes. If there is no similarity in the topological structure of the phylogenetic tree, then there is no related function. If the topological structure of phylogenetic tree is similar, then it has related function. We often call this kind of tree with similar structure mirror tree.

3.4 Application of Protein Structure Method

If different proteins interact with each other, it means that they have corresponding structural basis, which is mainly manifested in affinity, specificity and so on. Compared with other methods, protein structure method has some advantages. In the prediction of protein-protein interaction, the information obtained by this method is more accurate and precise. There are many methods to predict protein-protein interaction from the structural information of proteins, among which the two main methods are multi-body stringing and homologous modeling. Firstly, multi-body stringing takes known proteins as a basic skeleton. By inserting the predicted sequence into the skeleton, we can observe the folding of the sequence in the skeleton and analyze the possibility of folding. This method is very vivid, mainly referring to the way of Chinese medicine, taking the known structure of the protein as a reference to effectively analyze the prediction of protein folding type. By comparing the characteristics of the two, the problem of predicting protein structure is transformed into the problem of known protein structure, and the corresponding data information is found in the existing protein database. It should be mentioned here that the multi-body stringing method only needs to predict the primary structure and the tertiary structure, and does not need to predict the secondary structure. By matching the known protein structure with the unknown protein sequence, and then studying the similarity degree of the known protein sequence, we can get the structure information of a variety of unknown proteins, and find the most suitable structure in a variety of structure information for protein prediction. While locating the position of protein, we need to calculate the force between residues and hydrophobic interaction, find the most stable position and optimal energy, and calculate the workload. From the above analysis, this method is suitable for predicting the core structure of protein[3]. Multi-body string method assumes that the folding types of proteins are limited. Therefore, only when the structure of unknown protein is similar to that of known protein, can the structure of unknown protein be predicted more accurately. When predicting the protein type, if the structure of unknown protein has never appeared before, this method can’t be applied. The homology modeling method is to use the function information of known proteins to homologous proteins. If we regard prot eins as a family, there are many homologous protein sequences in this family, whose spatial structures are similar. When the protein sequences are the same, the homology modeling is successful.

4. Conclusion

To sum up, bioinformatics is a comprehensive subject. There are many methods to predict protein-protein interaction. How to combine different methods reasonably is an important subject of bioinformatics research.

References

[1] Wang Xia, Li Beiping, Tan Mingfeng, Wang Yuelan, Yue Junjie, Liang Long. Prediction of Functional Modules in Protein-Protein Interaction Network by Bioinformatics Method. Biotechnology Communication, vol.20, no.03, pp.430-432, 2009.

[2] Wang Wei. Establishment of High Throughput Yeast Two Hybrid Platform and Its Application in Proteomics. Fudan University, 2007.

[3] Liu Zhen. Research on Bioinformatics in Proteomics. Hunan Normal University, 2005.