Interaction-Based Analysis of Protein Function and Network

Haretsugu Hishigaki [1,2]
hisigaki@ims.u-tokyo.ac.jp

Toshihide Ono [2]
ono@otsuka.gr.jp

Akira Tanigami [2]
atanigam@otsuka.gr.jp

Toshihisa Takagi [1]
takagi@ims.u-tokyo.ac.jp

[1] Laboratory of Genome Database, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minatoku, Tokyo 108-8639, Japan
[2] Otsuka GEN Research Institute, OTSUKA Pharmaceutical Co., Ltd., 463-10 Kagasuno, Kawauchi-cho, Tokushima 771-0192, Japan

1 Introduction

Whole genome sequences of several organisms have been completely determined by advances of genome sequencing projects. Consequently, many of the novel gene and protein structures have been also identified. However, many of genes or proteins have not been annotated with biological functions. Thus the next major challenge of genome sequencing project are to characterize the biological function of each protein and to elucidate its roll in various intracellular processes, such as translation, splicing, post-translational modification, post-translational transfer, and so on [1]. For identification of gene functions and elucidation of intracellular processes, many protein-protein interactions, which are not only physical interactions between two proteins, but also functional or genetic interactions, have been rapidly identifying and their experimental results have been accumulating in the public databases such as MIPS and YPD [2, 3].
We propose the computer software. Two major functions of this software are:

1. inferring the protein functions from the relationships between the specified protein and the others, in other words, supporting the knowledge-based decision making for protein functions.

2. finding the plural protein networks between specified two proteins in the protein linkage map, which is constructed by integration of all protein-protein interactions.

Using this computer software, a user might be able to obtain several predicted annotations about gene functions and to discover plural possible protein networks that might include the novel and/or the alternative protein networks.

2 Method

1. Material
   We obtained 1126 protein-protein interactions which are physically interacted with each other. And 902 distinct proteins of yeast *Saccharomyces cerevisiae* are included in all protein-protein interactions. We obtained the "Sub_cellular localization catalogue" data, which contains the information about protein localization sites, that is, the information about where each yeast protein located in the living cell. And we also obtained the "Functional catalogue" data of yeast proteins. These data are obtained from MIPS database.

2. Graph representation of protein linkage map
   We adopted the graph to represent the protein linkage map by integration of large amounts of protein-protein interaction data. Each protein and physical connection between two proteins is denoted as each node and edge in a graph, respectively. Each node is defined three biological information as its attributes, such as the "description", the "synonym", and the "protein localization site" in the cell. The protein linkage map is represented as an undirected graph, because in this case of the protein-protein interactions, the connections between two proteins are not considered their directions.

3. Inference of the gene function
   Connection between an unknown and a well-known protein, as well as sequence homology to a known protein family, can provide a convenient shortcut for detecting a biological function of an unknown protein. For example, if the molecular biologists have no information about a protein, they will infer its biological function, such as the protin localization site, the cellular role of the protein, and so on, from the annotations of neighbor proteins in a protein linkage map. To achieve their requirement, we implemented the computer software to support the knowledge-based decision making for protein function.

4. Finding the plural protein networks in a protein linkage map
   For finding the plural protein networks between two proteins specified by a user, the "K-th shortest path algorithm" is adopted and implemented. This algorithm determine not only the shortest path which is defined as the minimum distance from one protein to another one, but also the second shortest, the third shortest, and so up to the K-th shortest path. See Martins et al. [4] for more details about this algorithm.

Acknowledgments

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, "Genome Science", from the Ministry of Education, Science, Sports and Culture in Japan.

References

[1] Karen, E., Andy, B., Norman, P., and Charlie, H., INTERACT: an object oriented protein-protein interaction database, *Proceeding of the ISMB '99*, 87–94, 1999.

[2] *MIPS database*, http://websvr.mips.biochem.mpg.de

[3] *YPD: Yeast Proteome Database*, http://www.proteome.com

[4] Martins, E.Q.V., Pascoal, M.M.B., and Santos, J.L.E., The K shortest paths problem, *unpublished*, http://www.mat.uc.pt/~eqvm/cientificos/investigacao/r_papers.html