Poster J-37

A Homology-based Method for Predicting Complexes in Large Protein-Protein Interaction Networks

Authors:
Jae-Hun Choi (*ETRI*)
Jong-Min Park (*ETRI*)
Soo-Jun Park (*ETRI*)

Short Abstract: In this paper, we propose a method for predicting biological complexes, in which interactions among their proteins are described specifically. The prediction suggests candidate complexes for a target species, which are homologically transformed from the complex experimented on well-known other species such as yeast, drosophila, mouse etc.

Long Abstract:

The networks for protein-protein interactions (PPI) can be represented by a set of relationships, which can model biological events among many of proteins in a cell. As an integrated data, it is used as an important mean to systematically understand major biological mechanisms taking place in a cell. In this network, there are many complexes that consist of several protein interactions to perform discrete biological functions. The complexes may be viewed as fundamental building blocks of cellular organization. Prominent examples of the blocks are “Hemoglobin,” “Ribosome”, “RNA Polymerase” etc. Especially, particular complexes related to some diseases may be a target for drug discoveries.

Currently, the PPI network data is quickly extracted by the enhanced experimental techniques, such as ‘Yeast Two-Hybrid’. Also, protein complexes are identified by high-throughput method such as ‘TAP-MS’ and ‘HMS-PCI’. However, the methods require much time and expense to detect specific user-interested complexes due to repeated experiments.

Therefore, alternative methods have been developed to predict target complexes before experiments and to detect known complexes for reviewing and supporting further analysis. Recently, Bader and Hogue suggest a clustering method for searching dense sub-networks in which each protein has similar functions with the others. This method relies on the fact that protein complexes generally correspond to dense sub-networks and its components have similar functions with each other. However, the method cannot find other complexes having different functions yet conceptually connected.

As another approach, Leser defines a query language, PQL(Pathway Query Language) for detecting user-intended complexes. This approach can detect relatively various structures by exploiting a set of constraints described in ‘WHERE’ clause, such as attributes of protein and path expressions between two proteins. However, it fails to automatically detect complexes, because the only way to describe the structure of the complexes is to manually formulate the query.

To resolve the problems, we propose a method for predicting biological complexes, in which interactions among their proteins are described specifically. The prediction suggests candidate complexes for a target species, which are homologically transformed from the complex experimented on well-known other species such as yeast, drosophila, mouse etc.

Additionally, the specific interactions among proteins involved in the complex are detected by mapping the proteins into PPI networks.

In general, the biological homology between two different species may exist for corresponding proteins and their interactions. Since the complex is represented as a set of PPIs, a complex of target species may also take the homology with one of other species. The following explains our homology-based method for predicting complexes. The prediction is divided into two processes; protein homology indexing and complex transformation.

The former makes a homology index for corresponding proteins between two different species, which biologically takes the similar features each other. In the latter, the proteins involved in complexes of model species can be transformed into corresponding proteins of target species by the homology indices. We can generate a virtual complex by detecting the interactions among them from PPI networks for model species. In sequence, as the virtual complex is mapped into a PPI network for target species, it is converted into the predicted complex. The employed data set is followed as Swiss-Prot for homology indexing among proteins, DIP for PPI networks and “yeast protein complex database” for model complexes respectively.

The homology between two proteins corresponding to each other species is evaluated by taking a comparison with their attribute values in protein database. The name, gene, GO term and amino-acid sequence of proteins are good examples as the comparison attributes. A pattern matching is employed to compare with protein names including their synonyms. The similarity between GO terms to express protein features is evaluated by conceptual distances of them located in GO hierarchies. In specific, we employ a BLAST tool to evaluate the similarity for two amino-acid sequences. We determine that two proteins are homologous, if the similarity values may exceed our threshold.

Since a complex in the database consists of only a list of one or more proteins, we should find the interaction relationships among proteins included in the complex. Fortunately, the relationships are easily detected by directly mapping their proteins into the PPI network of model species. Then, we generate a virtual complex including the interaction relationships among proteins, as the proteins in the complex of model species are converted into ones of target species by their homology indices. Since a protein takes homology with one or more proteins, our model enables users to optionally select all or the best one of them. We make a refinement of the relationships among proteins in the virtual complex by integrating them with the relationships extracted by mapping the proteins into the PPI network of target species. The refined complex is suggested to users as the predicted complex for the target species, which takes a homology with an input complex for the model species.

Reference

[1] A.C. Gavin, M. Bosche, et al., “Functional organization of the yeast proteome by systematic analysis of protein complexes”, Nature, 15(6868). 2002.

[2] G. D. Bader and C. W. Hogue, “An automated method for finding molecular complexes in large protein interaction networks,” BMC Bioinformatics., 4:2, 2003.

[3] U. Leser, “A query language for biological networks,” Bioinformatics, vol.21 Suppl. 2:ii33-ii39, 2005.

[4] T.W. Huang, "POINT: a database for the prediction of ppi based orthologous interactome", Bioinformatics, Vol. 20, No. 17, 2004.