A Method for Protein Functional Flow Configuration and Validation

Woo-Hyuk Jang[1] torajim@icu.ac.kr
Suk-Hoon Jung[1] jsh@icu.ac.kr
Dong-Soo Han[1] dshan@icu.ac.kr

[1] School of Engineering, Information and Communications University, 119, Munjiro, Yuseong-gu, Daejeon, 305-732, Korea

Abstract

With explosively growing PPI databases, the computational approach for a prediction and configuration of PPI network has been a big stream in the bioinformatics area. Recent researches gradually consider physicochemical properties of proteins and support high resolution results with integration of experimental results. With regard to current research trend, it is very close future to complete a PPI network configuration of each organism. However, direct applying the PPI network to real field is complicated problem because PPI network is only a set of co-expressive proteins or gene products, and its network link means simple physical binding rather than in-depth knowledge of biological process. In this paper, we suggest a protein functional flow model which is a directed network based on a protein functions' relation of signaling transduction pathway. The vertex of the suggested model is a molecular function annotated by gene ontology, and the relations among the vertexes are considered as edges. Thus, it is easy to trace a specific function's transition, and it can be a constraint to extract a meaningful sub-path from whole PPI network. To evaluate the model, 11 functional flow models of Homo sapiens were built from KEGG, and *Chronbach's alpha* values were measured (alpha=0.67). Among 1023 functional flows, 765 functional flows showed 0.6 or higher alpha values.

Keywords: signal transduction pathway, functional flow, *Chronbach's alpha*, protein, gene ontology

1 Introduction

At the early studies of PPI prediction, many prediction techniques were developed based mainly on a few features of a protein (i.e., domain frequency in the interaction protein pair), so they suffered from low prediction accuracy problem. However, recent researches[1,2] gradually consider physicochemical properties of proteins and support high resolution results with integration of experimental results. With regard to current research trend, it is very close future to complete a PPI network configuration of each organism. The signal transduction is a process which describes a cell change by external stimulus, and it plays an important reference roll of most fundamental cellular process. Most of signal transduction is initiated by extra cellular signal, and cascade intracellular activities by ligand-receptor binding are followed such as protein phosphorylation and de-phosphorylation, PPI, and protein-small molecules interaction. Given the fact that signal transduction pathway is protein's cascading activities, identifying participants and their relationship for a specific signal transduction from whole PPI network is an essential work. However, even though the PPI network is completely configured, extracting signal transduction pathway is complicated problem because PPI network is only a set of co-expressive proteins or gene products, and its network link means simple physical binding rather than in-depth knowledge of biological process. Thus, most of the target signal transductions have been manually discovered so far. To overcome the problem, we suggest a protein functional flow model which can be a constraint to extract meaningful sub paths from whole PPI network. Suggested model is a directed network based on a protein functions' relation of signaling transduction pathway.

2 Functional Flow

The functional flow model is a concept which extends protein-protein relationship to function-function one. In this research, protein and PPI information were gathered from known signaling transduction pathways. Protein function in this paper is a molecular function which is one of the categories of gene ontology (cellular component, biological process, molecular function). Figure 2-1 shows some proteins and their relationship of Hedgehog

signaling transduction pathway. Each protein has general molecular functions and the functions have relations such as inhibition or dissociation.

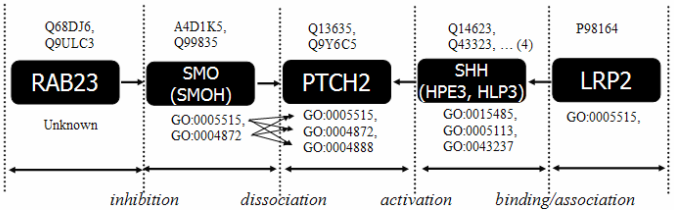

Figure 2-1. Some proteins and their relationship in Hedgehog Signaling Pathway

Based on the relation "binding/association" between "LRP2" and "SHH", we consider that protein binding(GO:0005515) , cholesterol binding(GO:0015485), patched binding(GO:0005113) and laminin-1 binding(GO:0043237) functions have "binding/association" relation. Proteins whose function is unknown were manually removed, and redundant count of functional flows was utilized as a weight score. Similarly, total 1023 functional flow were extracted from 11 *H. sapiens* signaling transduction pathways of KEGG database.

3 Validation

Internal integrity of functional flow was measured via *Chronbach's alpha* value. *Chronbach's alpha* value checks integrity or similarity of each questionnaire when single concept is asked by many different questionnaires. The variables of *Chronbach's alpha* values correspond to followings.

N = total count of functional flows which extracted from a specific signaling transduction pathway.

σ^2_i = a dispersion of a specific functional flows out of 11 signaling transduction pathways.

σ^2_j =a dispersion of all functional flows in a specific signaling transduction pathway.

Note that the type of a certain functional flow has conflict in other signaling transduction pathway, we decrease a appearance values from 11 to zero. The average alpha of overall functional flows was 0.67, and 765 functional flows had 0.6 or higher alpha values. Table 3-1 and Figure 3-1 shows details.

Table 3-1. Validation result for each functional flow

ID	tFR	Avg. freq.	Avg. dispersion	Avg. alpha	a>0.6
04010	255	9.43	7.56	0.63	211
04012	90	9.76	6.03	0.72	78
04310	118	8.41	11.29	0.48	36
04330	28	8.68	8.50	0.60	17
04340	29	8.31	14.15	0.54	10
04350	85	9.27	8.46	0.66	73
04370	89	8.96	6.31	0.64	50
04630	154	9.33	6.03	0.65	75
04020	51	10.24	7.24	0.86	47
04070	186	10.76	1.07	0.87	169
04150	52	9.56	7.05	0.71	46

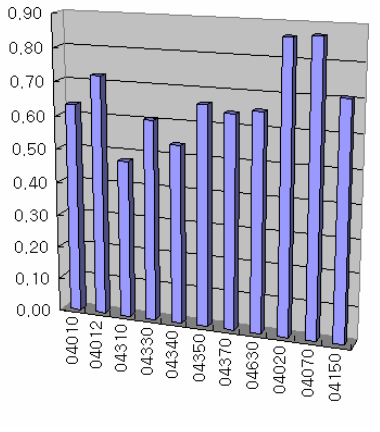

Figure 3-1. *Chronbach's alpha* distributions.

Acknowledgement

This research was financially supported by the Ministry of Education, Science Technology (MEST) and Korea Industrial Technology Foundation (KOTEF) through the Human Recourse Training Project For Regional Innovation.

References

[1] Ali Cakmak and Gultekin Ozsoyoglu, "Mining biological networks for unknown pathways", *BIOINFORMATICS*, 23:20, 2007.

[2] Amy Hin Yan Tong, "Global Mapping of the Yeast Genetic Interaction Network", *Science*, 303:5659, 2004.