The Construction of the Knowledge Base of Immune System

Yoshinobu Igarashi igarashi@kuicr.kyoto-u.ac.jp
Minoru Kanehisa kanehisa@kuicr.kyoto-u.ac.jp

Institute for Chemical Research, Kyoto University
Gokasho, Uji, Kyoto 611-0011, Japan

1 Introduction

The paradigm shift from experimentals to bioinformatics is apparent in many areas of biology. There are immediate demands to construct databases that are designed for computer analysis of biological phenomena. Specifically, the immune system has been researched very intensely in experimentals because it is essential to protect our bodies against viruses, bacteria, and other foreign things. The immune system is made up of cell-cell interactions and cytoplasmic signal transductions. In recent years many molecules have been identified and characterized, as well as many interactions among them. Thus, we focus on the immune system and investigate most specified and embodied cell-cell interactions. Among multicellular organisms, only vertebrates can have immune system in their bodies, and various events would occur by complicated interactions between specified cells, T-cells, B-cells, macrophages, etc. In addition, cell differentiation is deeply crosslinked to the immune system in terms of cells' roles, because cells are changing their characters through differentiation by ligand stimulation. Therefore when we try to understand the immune system, we must refer to cell differentiation in terms of signal transductions and analyze the pathways that result from many interactions. However, there is virtually no database for the immune system that refers to signal transductions and cell-cell interactions. Our database is designed in a similar way as BRITE [1] database which contains cell cycles and early developing stages, and we can analyze systematically the immune system, for example, by using the information of the whole genomes sequences.

2 Method

2.1 Knowledge base

We collected interaction data from experimental papers, and constructed a knowledge base of the immune system. The knowledge base is made as a flat file and is based on BRITE [1] format. It includes cell-cell interactions concerning CD molecules and signal transductions mainly involving cytokines. Our knowledge base is not yet made available as a WWW service, but we will incorporate these data into BRITE and open it to public.

2.2 Genome data

The analysis is made for the whole genome sequences of 17 species:

Archaea	*M. jannaschii*	*M.thermoautotrophicum*	*A. fulgidus*
	P. horikoshii		
Bacteria	*E. coli*	*H. influenzae*	*H. pylori*
	B. subtilis	*M. genitalium*	*M. pneumoniae*
	M. tuberculosis	*A. aeolicus*	*Synechocystis*
	C. trachomatis	*B. burgdorferi*	*T. pallidum*
Eukaryotes	*S. cerevisiae*		

In addition, 11 species with partial genomes are analyzed. All genome sequences are taken from GENES database in KEGG [2].

2.3 Sequence similarity search

Firstly, we detected homologues of the molecules in our database. The similarity search was performed by ssearch [3] with cutoff value of 250. Secondly, we evaluated the alignment and additional information with manual checking. This includes the position in the genome and we also checked the neighbors.

3 Results and Discussions

We have found three sequences which have high similarity with human integrin alpha chains in *Synechocystis*, though *Synechocystis* is not a multicellular organism. We also found WD repeat sequences, which are known to be found in only eukaryotes, in the neighbour upstream of the integrin alpha-like sequences in *Synechocystis* [4]. Therefore we guess that there is a possibility that these sequences are not original in *Synechocystis* but transferred from eukaryotic organisms. Furthermore, *Synechocystis* has homologs of cadherins that are engaged in cell-cell adhesion in multicellular organisms. We will analyze and detect molecules that are interacting with these homologs.

Acknowledgements

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 of Japan. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References

[1] Kawashima. S., Katayama, T., and Kanehisa, M., Construction of Molecular Interaction Database and Searching for Similar Pathways *Genome Informatics 1997*, Universal Academy Press, 298–299, 1997.

[2] Ogata. H., Goto, S., Fujibuchi, W., and Kanehisa, M., Computation with the KEGG pathway database, *Biosystems*, 47:119–28, 1998.

[3] Pearson. W.R., Effective protein sequence comparison, *Methods Enzymol.*, 226:227–58, 1996.

[4] Kaneko. T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S., Sequence analysis of the genome of the unicellular *cyanobacterium Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions, *DNA Res.*, 3:109–36, 1996.