Developing Molecular Interaction Database and Searching for Similar Pathways

Shuichi Kawashima, Toshiaki Katayama and Minoru Kanehisa

We have developed a database named BRITE, which contains knowledge of interacting molecules and/or genes concering cell cycle and early development. Here, we report an overview of the database and the method of automatic search for functionally common sub-pathways between two biological pathways in BRITE.

keywords: Knowledge-base/ Cell cycle/ Development/ Graph theory/ WWW

All known biological processes in living organisms are maintained and carried out by various molecular/ genetic interactions. While a large amount of knowledge on metabolic and regulatory pathways that has been accumulated is quite valuable in theoretical or computational studies for understanding biological systems, the knowledge mainly exists in research papers or reviews which are unsuitable for computational analyses. Besides, the existing molecular biology databases such as GenBank, EMBL, SWISS-PROT, and PDB, focus on representing the information about DNA, RNA or protein molecules and do not fully represent molecular interactions. Thus, we have initiated efforts to collect the knowledge in a different format and construct a new database of molecular interactions. In view of the surprising functional similarities that have been found in pathways among various species, the knowledge-base such as BRITE would lead to a new research area for analysis of biological networks.

The knowledge-bases concerning metabolic pathways have already been developed, such as KEGG, EcoCyc, and WIT, and proved to play important roles in bioinformatics. We are constructing a new knowledge-base named BRITE [1] (Biomolecular Relations in Information Transmission and Expression) which focuses on regulatory pathways in various biological processes. BRITE is a flat-file collection of entries, where each entry contains a relation of two molecules. The data representation of molecular relations in BRITE is based on KEGG [2], where the concept of 'biological links' is represented as follows:

organism:gene1 -> organism:gene2

This scheme can also be regarded as the following:

organism:relation

This is the basic component of an entry in BRITE.

MOLECULAR BIOLOGY AND INFORMATION — Biological Information Science —

Scope of research

This laboratory aims at developing theoretical frameworks for understanding the information flow in biological systems in terms of genes, gene products, other biomolecules, and their interactions. Towerd that end of a new deductive database is being organized for known molecular and genetic pathways in living organisms, and computational technologies are being developed for retrieval, inference and analysis. Other studies include: functional and structural prediction of proteins from sequence information and development of sequnece analysis tools.

Prof
KANEHISA, Minoru
(D Sc)

Instr
GOTO, Susumu
(D Eng)

Instr
OGATA, Hiroyuki

Students:
TOMII, Kentaro (DC)
SUZUKI, Kenji (DC)
KIHARA, Daisuke (DC)
KAWASHIMA, Shuichi (DC)
PARK, Keun-joon (DC)
HATTORI, Masahiro (DC)
BONO, Hidemasa (DC)
IGARASHI, Yoshinobu (MC)
KATAYAMA, Toshiaki (MC)
TAKAZAWA, Fumi (MC)
TANIGUCHI, Takeaki (MC)
NAKAO, Mitsuteru (MC)

Research Fellow:
SATO, Kazushige (RF)
SUGIYAMA, Yukiteru (D Agr)

Each entry in the BRITE database contains the ENTRY and DEFINITION lines at the top. The ENTRY line holds a unique identifier and the summary of each entry is described in the DEFINITION line. The core section is RELATION, which describes the molecular interaction in the form of a binary relation. The RELATION section consists of three items: FROM for the regulating molecule, TO for the regulated molecule, and MESSAGE for the relation between two molecules, such as 'transcriptionally activate/repress', 'phosphorylate', and 'bind'. BRITE does not store detailed information of molecules that appear in the entry, but provides links to other existing databases for such information, including GENES, GenBank, and SWISS-PROT. The cross-reference information to other databases is described in the FACTORS section. The bibliographic information of references associated with the entry is described in the REFERENCE section with links to Medline.

Originally, molecular interaction data concerning only cell cycle controls were accumulated in BRITE. Now we collect data concerning developmental pathways as well. At present BRITE contains data about cell cycles of *S. cerevisiae*, *S. pombe* and *H. sapiens* and early developments of *D. melanogaster* and *X. laevis*. The amount of data in BRITE is shown in Table 1.

The BRITE database is available through the GenomeNet WWW service at the following address:

http://www.genome.ad.jp/brite/brite.html

We provide two facilities for data retrieval. One is retrieving an entry through the clickable map. The user can easily retrieve entries by clicking on a molecule or an interaction on the graphical pathway maps implemented in WWW. The other is by keyword search using the DBGET system. DBGET is the integrated database retrieval system with two basic commands, bfind and bget, to search and extract entries from a wide range of molecular biology database. The keyword search using bfind command is made against the DEFINITION line in BRITE. The result of the search is a list of entry names found, from which an entry can be selected to view the content. When the user already knows the entry identifier, it can be retrieved simply by using bget command. If the user has the amino acid sequence data for the molecular interaction of interest, a homology search can be performed against the BRITE database using the sequence as a query. This is especially useful when searching similar interactions and similar pathways in different species or in different biological processes.

Table 1. The number of entrys in BRITE

	Organism	Number of entries
Cell cycle	*S. cerevisiae*	64
	S. pombe	65
	H. sapiens	102
Development	*D. melanogaster*	80
	X. laevis	15

In order to automatically find functionally similar sub-pathways in the maps of BRITE we have developed a method similar to the one utilized for recognizing common structural fragments among chemical compounds [3]. By regarding each molecule and relation in a map as a vertex and edge, respectively, a pathway in BRITE is represented as a graph. Because the method introduces relations among all molecules irrespective of whether they exist or not in the database, the graph representing a pathway is defined as an edge-weighted complete graph. Then we construct a docking graph from two edge-weighted complete graphs. Searching for an edge-weighted maximal common subgraph for two complete graphs is equivalent to searching for a clique (complete subgraph of graph) in this docking graph.

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

References

1. Tsukamoto N and Kanehisa M, Proc. Genome Informatics Workshop 1995, 158-159 (1995).
2. Kanehisa M, Trends Biochem. Sci., 22, 442-444 (1997)
3. Takahashi Y, Maeda S and Sasaki S, Anal. Chim. Acta, 200, 363-377 (1987).