BRITE: Biomolecular Reactions for Information Transmission and Expression

Susumu Goto, Kenji Suzuki, Yutaka Akiyama and Minoru Kanehisa

We are developing a knowledge base of "Biomolecular Reactions for Information Transmission and Expression" (BRITE). As a first step of this project, we developed a signal transduction database and a metabolic pathway knowledge base. The signal transduction database and the metabolic pathway knowledge base represent molecular interactions involved in the signaling pathways and enzyme interactions, respectively. Both are linked to the other public databases such as PIR protein sequence database, the PDB protein three-dimensional structural database, and the OMIM database on genetic diseases. We provide a graphical user interface to access them.

Keywords: Biomolecular reactions/ Molecular interaction/ Signaling pathways/ Metabolic pathways/ Database/ Knowledge base/ Graphical user interface

The signal transduction from extracellular signals to gene expression is one of the significant cellular events that are beginning to be unraveled in molecular details. The event starts at the cell surface receptor accepting an external signal. The signal is transmitted inside the cell to key signaling molecules such as Ras and G-protein. The activation of these molecules is often followed by a cascade of protein phosphorylation events which results in the activation of specific transcription factors in the cell nucleus. Such an overall picture is derived from numerous experiments on molecular interactions, i.e., data on one molecule affecting other molecules either directly or indirectly.

There are various kinds of molecular biology databases specialized to amino acid sequences, nucleotide sequences, protein three-dimensional structures, and so on. However, they are not the databases for overall picture on molecular interactions, but for one molecule or a piece of data. So far, the picture is only in the biologists' image. Our long-term objective is to automatically construct such an overall picture from pieces of data stored in molecular interaction databases.

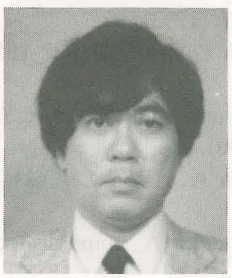
Toward that end, we have started collecting data on molecular interactions that play parts in the signal transduction pathways and experimenting various representations and manipulations of those data. Table 1 shows one possible description of molecular interactions in the Ras pathway, where the basic element is a pair of interacting molecules or molecular complexes, called a donor and an acceptor, together with the description of molecular events [1].

We have also started developing a method to automatically construct the pathways and their graphical views from pieces of data. We exploited data on

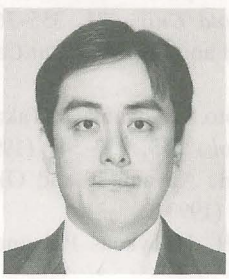
MOLECULAR BIOLOGY AND INFORMATION —Biological Information Science—

Scope of research

The following five attempts have been mainly made in this laboratory. (1) Characterization of amino acid sequences by extracting signature oligopeptides from protein structure and sequence databases. (2) Characterization of nucleotide sequences around promoter, translation initiation and splice sites. (3) Construction of new databases that describe molecular interactions, such as signal transduction and metabolic pathways. (4) Modeling three-dimensional structure of RNA, DNA and protein. (5) Development of database systems and tools to support researchers in genome community. Almost all of them can be used to access the databases maintained in this laboratory and analyze data from all over the world via Internet.

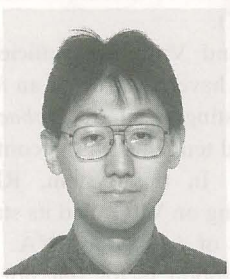


Professor
KANEHISA, Minoru
(D Sc)



Associate Professor
AKIYAMA, Yutaka
(D Eng)

Instructor
UCHIYAMA, Ikuo



Instructor
GOTO, Susumu
(D Eng)

Students:
MIZUNO, Masahiko (DC)
OHKUBO, Zenmei (DC)
FUJIBUCHI, Wataru (DC)
OGATA, Hiroyuki (DC)
SUZUKI, Kenji (MC)
TOMII, Kentarou (MC)
YOSHIDA, Ayahiko (MC)
KIHARA, Daisuke (MC)
TSUKAMOTO, Nobuo (MC)
FUKUSHIMA, Nobuhiro (RF, D Sc)
PARK, Keun-joon (RS)

metabolic pathways, which is another data on molecular interactions and is better known than the signal transduction, as an application. Using the enzyme reaction database LIGAND, we developed a knowledge base system for searching and browsing metabolic pathways [2]. Biologists can ask the information about metabolism like below with the knowledge base,

- whether there is an alternative pathway in case that the given enzyme is deficient or altered, and
- which enzyme or pathway relates to the given disease.

Because the system constructs enzyme networks only by traversing the common substrates in two reactions, improvement of this system is indispensable to construct more precise view of the interactions. For example, interface and mechanism to use various condition such as organism and site specific information and quantitative information on the substrates can improve the system.

Another and our short-term objective is to provide a browser of different types of data in an integrated environment. Thus, the molecular interaction data are linked to consensus views of the transduction pathways or metabolic pathways, as well as to other databases including Medline (bibliographic data), SWISS-PROT, PIR (amino acid sequence data), PDB (protein three-dimensional structure data), LIGAND (chemical compounds in enzyme reactions), and OMIM (genetic diseases). Using the links, users can easily retrieve the related information stored in the separated databases. We also developed the LinkDB that links databases not only directly but also indirectly or reversely. Some databases have a lot of cross reference information, but others have few. If users retrieve the database that has few cross reference, they have to repeatedly retrieve other databases until they have required information. The LinkDB provides the precomputed indirect and reverse links, thus users can retrieve necessary cross reference information directly.

There are two alternative ways to provide user interface to access the databases. One is the use of Mosaic in the World Wide Web (WWW) system. Mosaic has an easy-to-use graphical user interface and we have already developed an integrated retrieval system called WebDBget linking sixteen databases in molecular biology. We have also provided LinkDB as a part of WebDBget. The other way is to provide graphical interface by our own software. Since the Mosaic interface does not provide sufficient functions to dynamically draw pictures of pathways for now, we adopted our own interface for metabolic pathway knowledge base to draw dynamically constructed pathways. We are consulting which interface is better for the BRITE system.

Table 1. A description of molecular interactions in the Ras pathway. The basic element is a pair of interacting molecules or molecular complexes, called a donor and an acceptor, together with the description of molecular events. The description of molecular events is not shown, but like "GF binding RTK leads to RTK dimerization & autophosphorylation" for the first line.

Donor	Acceptor	Signal	Interaction
GF	RTK	+	binding
RTK	GRB2	+	binding
GRB2	Sos	+	complex
GRB2/Sos	Ras	+	binding
Ras	Raf	+	translocation
Raf	MEK	+	phosphorylation
MEK	MAPK	+	phosphorylation
MAPK	Myc	+	phosphorylation
Myc	DNA	+	binding
MAPK	Jun	+	phosphorylation
Jun	Fos	+	dimerization
Jun/Fos	DNA	+	binding

Since this work is the first step to construct BRITE system, we have many works to do. One of the most important works is the automatic and intelligent construction of pathways from the pieces of data. We have an experience to construct long nucleotide sequence from the pieces of sequences under various conditions by using techniques in deductive database systems [3]. Those techniques can be also used to construct pathways and we are planning to implement them as a part of the BRITE system.

Acknowledgment

This work was supported by the Grant-in-Aid for Scientific Research on the Priority Area 'Genome Informatics' from the Ministry of Education, Science and Culture of Japan.

References

1. Suzuki K, Goto S, Akiyama Y and Kanehisa M, *Proc. Genome Informatics Workshop 1994,* 144–145 (1994).
2. Goto S and Kanehisa M, *Proc. Genome Informatics Workshop 1994,* 234–235 (1994).
3. Sakamoto N, Goto S and Takagi T, *Intl. J. Bio-Medical Computing,* **36**, 171–179 (1994).