A Data and Knowledge Base for Cell Signaling Networks

Takako Igarashi[1] Yoko Nadaoka[2] Tsuguchika Kaminuma[1]
taka@nihs.go.jp nadaoka@rinshoken.or.jp kaminuma@nihs.go.jp

[1] National Institute of Health Sciences
1-18-1 Kami-Yoga, Setagaya-ku,Tokyo 158
[2] The Tokyo Metropolitan Institute of Medical Science
3-18-22 Hon-Komagome, Bunkyo-ku, Tokyo 113

Abstract

Each cell in a multicellular animal is programmed during development to respond to specific outer chemical signals. Some of these chemical signals activate receptor proteins on the surface of the cell that trigger series of membrane and intracellular signal transductions, and eventually influence gene expressions. These complex cell signaling mechanisms have been unveiled at molecular levels in various multicellular organisms in the past decade. It was found that these molecular signaling pathways or what we may call cell signaling networks (CSN) play important roles in wide range of biological phenomena that characterize multicellular animals. These phenomena include development, differentiation, reproduction, morphogenesis, carcinogenesis, apoptosis, and even learning.

We have developed the data and knowledge base of the CSN that consists of interacting extracellular (xenobiotic) chemicals and biomolecules. The system contains signaling pathways, and structural and functional data of the molecules. The system was implemented on UNIX workstations using an object oriented database management system ACEDB.

1 Introduction

Each cell in a multicellular animal is receiving chemical signals. Some of these signals are from its neighbors (or even itself) and some are from remote tissues. Hormones and cytokines are well known examples of such chemical signals. Some chemical signals come also from outside of the animal. Drugs and toxic substances are such examples. Cells have receptors both for

[1]五十嵐貴子, 神沼二真: 国立衛生試験所, 東京都世田谷区上用賀 1-18-1
[2]灘岡陽子: 東京都臨床医学総合研究所, 東京都文京区本駒込 3-18-22

biological chemical signals and for xenobiotic chemical signals. In either case, some of these signal molecules bind to receptors on the cell surface, while some smaller signal molecules penetrate into the plasma membrane and directly bind to intracellular signal molecules[1,2].

By analogy each step of cell signal transfer can be considered as a switch, a gate, or a transistor in electric circuits in LSI (Large Scale Integrated Circuit), and the whole set of such signal pathways forms a very complex and sophisticated information processing system[3]. As cells are building blocks for living organisms, so the networks are units for the information systems in the entire organism.

Before 1980 few such cell signal molecules were known. In 1980s however many signal molecules had been discovered. In 1990s more signal molecules have been found and even three dimensional structures of some of these important signal molecules were determined.

Even though the entire set of large and complex cell signaling networks has not yet completely unveiled, quite a number of characteristic features of such signal molecules and networks have already been unveiled. Moreover the pace of accumulation of data and knowledge in this research field is ever increasing so that it now becomes very difficult for an individual researcher to comprehend and correlate enormous amount of the existing information.

About two years ago we started to develop a computerized data and knowledge base for cell signaling networks (Cell Signaling Networks Database : CSNDB). The eventual goal of this project is to catalog all networks and their signal molecules in order to model cells as information processing units, and to make more realistic and quantitative models for development, differentiation, morphogenesis, carcinogenesis, neural information processing, and learning. In this paper we shall introduce the first prototype of the system.

2 Methods

2.1 System Implementation

The CSNDB was implemented on an object oriented database management system, ACEDB[4]. In ACEDB, the data are organized into objects and classes. Objects are knowledge items and classes are sets of similar objects. These similar objects are represented by a similar hierarchical tree model. A tree model contains numbers, sentences, and pointers to other objects. Pointers are represented by names for other relevant classes. Thus the objects are linked by these pointers, and they can be browsed freely by mouse clicking like a hypertext. In the case of numbers or sentences their data types are specified by such code words as Int, Float, Text, and Greek.

ACEDB runs on X-window environment on UNIX workstations. Using ACEDB, we have implemented the CSNDB on our SUN Sparc workstations and Silicon Graphics, Indigo and, Indy workstations.

In general an object oriented database is flexible for changing its data structure, but this strength often makes users difficult to comprehend relations among data items and to comprehend over all data structure. In order to overcome this weakness ACEDB provides search function module. The data search is carried out by user typed commands that consist of standard "and", "or", and "not" operation.

ACEDB also has the function to call external programs. This function enable CSNDB to handle not only numbers and sentences but also graphical and dynamical data. Using

this function we have linked other programs to CSNDB system. These include the program generating diagrams (Metab Display [5]), molecular viewer program (RasMol [6]), picture viewer program (XV [7]), Internet browser (Netscape [8]), and help function. Help system illustrates how to use this system. These help information are written in html format, and can be viewed by (internal) the html browser. When necessary the user click the help menu in a window to call this program.

Since the ACEDB contains a WWW interface [9], CSNDB implemented on ACEDB can be accessed through the Internet.

2.2 Data Sources

Knowledge on cell signaling networks and signal molecules were sorted out from standard text books, reviews, and original papers [10-15]. In addition, we used computerized databases for collecting facts data of chemicals and biomolecules. We referred to GenBank for DNA and protein sequence, to Protein Data Bank for three dimensional structures of protein, to the Cambridge Structural Database for three dimensional structure of chemicals, and to MEDLINE for literature.

3 Results

3.1 Information Contents

The Cell Signaling Networks Database (CSNDB) system was designed to provide information on Signal Pathways and Signal Molecules. The information of Signal Pathways includes the flow of signal transduction and the information of Signal Molecules includes the characters of signal molecules which are the components of signal pathways. Contents of these two categories of information are the following;

- Signal Pathways : pathway flow diagrams, and reactions to transfer signals.
- Signal Molecules : hormones, cytokines, receptors, ion channels, effectors, enzymes, messengers, and transcriptional factors.

  Each signal molecule is characterized by the following biological attributes;

  1. Genomic Data : exon and intron structures of human genome, DNA sequences, and protein sequences.
  2. Structural Data : domain structures, functional motifs, 3-D structures, and ligand bindings.
  3. Mutational Data : mutational sites, functional changes, and relating diseases.
  4. Associated Chemicals : activators, inhibitors, agonists, antagonists, toxins, and drugs.
  5. Bibliographical Data : papers, and the date of the last update.

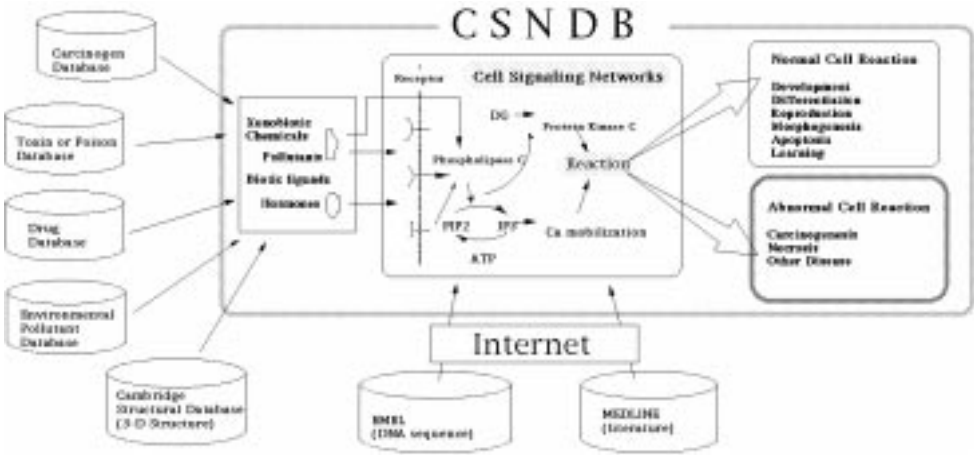

Figure 1: Concept of Cell Signaling Networks Database.

Since the CSNDB is implemented on the object oriented database, all these categorized information are defined as classes. Each class comprises individual data which are called objects. Each object has connection with other appropriate object one another. Windows of each object are opened by the clicking mouse.

For the present system, only human data for the cell signaling networks were compiled. This is because we put high priority on those cell responses that relate to certain important human diseases such as cancer and aging that may be resulted from mutation and/or other biological changes of signaling molecules. We are planning to expand the system to include data on other species such as Drosophila, Caenorhabditis elegans, and Escherichia coli in the next step.

The present CSNDB contains the following characteristic signal pathways: growth factors, steroid super family, cAMP and PKA, IP3 and PKC, adaptor molecules, ras / raf and MAPK, cytokine, and JAK / STAT. We have stored 53 signal molecules, 837 sequences, 67 chemicals, 68 3-D structures, 147 pictures, and 62 papers.

Figure 1 shows the concept of CSNDB. The eventual goal of the system is to simulate responses of a cell to chemical agents which have their origin either in the living system or outside. The key idea is that such a simulation system must be based on sound knowledge on inter molecular interactions among xenobiotic and biotic ligands and their complementary signal molecules in the cell. We postulate that such a simulation system should give good insight into many important multicellular biological phenomena such as development, carcinogenesis, apoptosis, and learning.

3.2 Representation of Signal Pathways

A typical signal pathway starts from extracellular stimuli and ends in physiological responses. What is interesting is that though there are wide variety of specific receptors corresponding to chemical ligands, there exist fewer common reaction pathways and their associated signal molecules that transfer the signals. The external stimuli are perceived by cell membrane receptor, transformed into intracellular signals, amplified, propagated, integrated, and at last

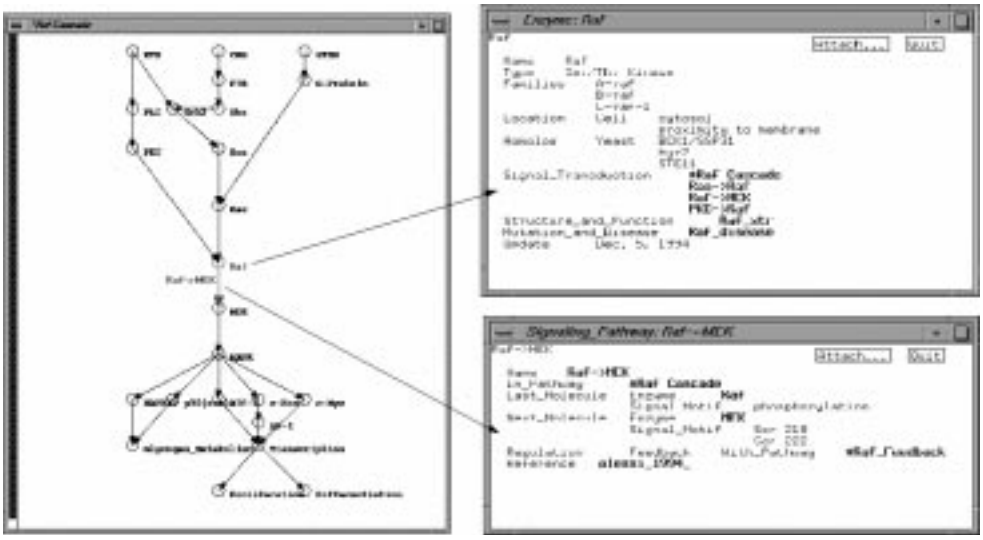

Figure 2: Examples of signal pathway diagram (left) and its corresponding deeper information (right).

trigger biological responses. Many biological responses are considered to be built not by a simple one way signal flow but by interactions among multiple ways. Some of these interactions are crosstalks which modify the direction of signal flow, some are feedback regulations which diminish or turn off the flow of the signal. As a whole signal transductions pathways compose complex networks.

In the CSNDB the signaling pathway motifs are represented as flow diagrams that consist of nodes and arrows (Figure 2). A node represents a signaling molecule and an arrow represents a molecular interaction that transduces the signal. Since the CSNDB has a hypertext structure, there exist hypertext links between flow diagrams and molecular information. If one clicks an arrow, the detail modes of molecular interactions are presented as a new window, and if one clicks a node, then the next window that contains deeper information about this signal molecule is displayed on the screen. Information of crosstalk and feedback are also embedded in the arrow objects and represented as flow diagrams which can be viewed when one clicks and opens the embedded information.

3.3 Representation of Signal Molecule

Protein Kinase C (PKC) is one of the most important signal transduction proteins, and enormous amount of experimental data have already been accumulated for this protein. We therefore take PKC an example to illustrate the kind information compiled for signal molecules.

Protein Kinase C (PKC) belongs to the class "Enzyme", and appears in the signal pathway "protein_kinase_C path". One can search detailed information on this protein by browsing the linked object data; synonyms, enzymatic types, families (isozymes), the gene site on the genome, exon and intron structure, nucleic acid sequence of the gene, homologs to other species, amino acid sequence of the protein, 3-D structure, domains and functional motifs, the phosphorylation

sites, activators, inhibitors, ligand binding sites, mutation sites and its effects, references, and date of the last update (Figure 3). One can click some of the items in the window to open further the next window in order to get more information about the item.

Domains and functional motifs of the protein are represented by schematic diagrams which is viewed by a drawing software XV. The three-dimensional structures of proteins and chemicals are represented by a molecular graphics software RasMol. One can rotate or magnify these graphical images in real time using the mouse. When necessary source data are fetched directory from GenBank and MEDLINE through the Internet; the Human Genome Center in Japan and National Center for Biotechnology Information in U.S.A., respectively. This connection is build by the browser, Netscape. XV, Rasmol, and Netscape are called from the CSNDB by request.

4 Discussion

Although system is still in its embryonic stage, we have already learned that our approach is reasonable and ACEDB is very suitable as the base system for implementing the CSNDB. In particular the flexibility of the object oriented database suits for integration of knowledge of molecular biology which may be adjusted or even will drastically be revised when new experimental facts are discovered. ACEDB also enabled us to handle diagrams, pictures, and 3-D images. These functions allow us to consider the CSNDB as a hypermedia database.

Right now the Internet (WWW server) version of CSNDB dose not support its full function, but this limitation will be taken away soon. We are also under construction of the more sophisticated WWW interface which can access other database such as Sybase and Oracle implemented on remote computers.

Though the CSNDB was originally designed to be a basic and common tool for wide range of researchers, it could also be used as a personal system by experimentalist who is studying specific receptors, cell signal transduction and other proteins. He may add his own data to the system or fill in his own data to the empty CSNDB.

5 Conclusion

We have developed the prototype of the data and knowledge base for cell signaling networks. We are planning to put it on our Internet WWW server for open usage. Once the system gets into that phase, we may expect to obtain comments and new data from other researchers through the Internet.

6 Acknowledgments

The authors thank to Dr. J. Thierry-Mieg and Dr. R. Durbin who started the ACEDB project and are very helpful for our work. The authors also thank to their colleagues ; Ms. M. Yukawa and Ms. S. Hasegawa who inputted PKC inhibitor data and drug data respectively, Drs. T. Nakano and K.Nakata who provided the Internet environment and the chemical and biological molecular databases. This work was partly supported by Japanese Human Science Foundation

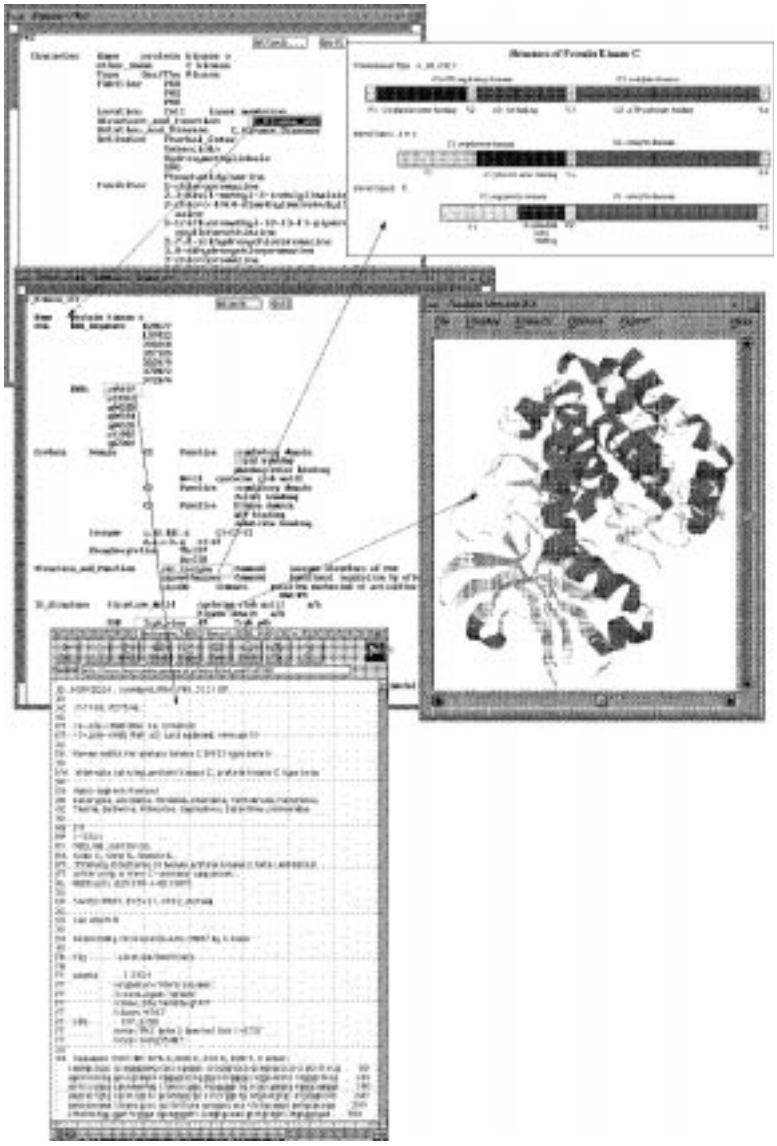

Figure 3: Data and knowledge items are cross referenced in windows.

Research Aid Program and by Science Research Promotion Fund from Science and Technology Agency.

7 References

1. Alberts, B. et al, Molecular Biology of The Cell (3rd ed.), Garland Pub. Inc., NY, 1994, Chap.15
2. Kandel, E. R. et al, Primciples of Neural Science (3rd ed.), Elsevier, Amsterdam, 1992, Chap12
3. Hunter, T., Cell, Vol.50, 1987. pp. 823-829
4. The softwares and documents of ACEDB and related topics were obtained from the following Internet sites:
   (a) Thierry-Mieg, J. and Durbin, R. acedb : A C.elegans Database: I. Users Guide., II. Installation Guide., III. Configuration Guide.
   ftp://ncbi.nlm.nih.gov/repositorv/acedb, 1991.
   (b) Gaeta, B. Tutorial Introduction to ACeDB.
   ftp://ncbi.nlm.nih.gov/repository/acedb/ace2, 1993.
   (c) Cartinhour, S. Exploring Models Design and Structure.
   http://probe.nalusda.gov:8000/acedocs/exploring/toc_models.html, 1994.
   (d) O'Callaghan, M. A Guide to Models and ace Files.
   http://probe.nalusda.gov:8000/acedocs/models_guide.html, 1993.
   (e) Kirk, J. How to install Guy Ducoux's WWW-ACEDB server.
   http://keck.tamu.edu/staff/ace_mosaic_howto.html.
5. Metab display is a software which runs in ACEDB environment developed by Dr. Stan Letovsky of GDB group of Johns Hopkins University.
6. XV can be obtained from ftp.cis.upenn.edu.
7. RasMol can be obtained from the following anonymous ftp site: colonsay.dcs.ed.ac.uk.
8. Netscape is a free software product from Communication Corporation.
9. WWW interface of ACEDB has been developed by Drs. Doug Bigwood and John Barnett of USDA National Agricalutural Library.
10. Krebs,E. G. et al. Trends Biochem. Sci. 19, E. G. 1994. pp.439-518
11. Marshall, C. J. et al. Cell 80, 1995. pp.179-278
12. Pennington, S. R.et al. PROTEIN PROFILE 1, 1994. pp.169-234
13. Vanhoutte, P.M.et al. Pharmacological Review 44, 1992. pp. 121-229, 349-457
14. Edited by Fleeger, C. A. USAN and the USP dictionary of drug names. MD, USA: United States Pharmacopoeial Convention, Inc. 1993.
15. Edited by Society of Japanese Pharmacopoeia. JAN: Japanese accepted names for pharmaceuticals. Tokyo, JAPAN: TAKUJI NIPPO, Ltd. 1992, and 1993.