Development of a Cell Signaling Networks Database

TAKAKO IGARASHI and TSUGUCHIKA KAMINUMA
Division of Chem-Bio Informatics, National Institute of Health Sciences, 1- 18-1 Kami-Yoga, Setagaya-ku, Tokyo 158, JAPAN

Abstract

In multicellular organisms cell signaling networks play important roles in wide range of biological phenomena, such as development, differentiation, reproduction, morphogenesis, carcinogenesis, apoptosis, and even learning. In order to explain biological phenomena based on cell signaling models, we have developed a database for cell signaling networks. The system contains mechanisms of signal transduction and structure and functional data and references of extracellular chemicals and biomolecules. CSNDB is constructed using ACEDB system, and includes various graphical representations such as pathway diagrams, map diagrams, 3-D images, pictures, and VRML environment. The system will be useful for modeling cells and their information processing, and to explain important biological phenomena based on these models.

1 Introduction

In living organisms biological reactions occur in order producing cascades, or biological pathways. There are two major pathways in cells of multicellular organisms; metabolic pathways and cell signaling pathways. Metabolic pathways had been studied by biochemists before the rise of molecular biology, and detailed knowledge of pathways had been accumulated and depicted on paper. Biochemical pathways provided from Boehringer Mannheim Co.[1] is a such example. Computerized databases for metabolic pathways have also been developed. EcoCyc[2] and Soybase[3] are such examples.

In 1980s however, the cell signaling mechanisms have been unveiled at molecular level in various multicellular organisms. It was found that cell signaling networks 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.

Since two years ago, we have started to develop a database for cell signaling networks (CSNDB)[4,5]. The eventual goal of this project is to catalog all the cell signaling networks in living organisms in order to make more realistic and quantitative models for biological phenomena. The system will be useful for modeling cells and their information processing,

and to explain important biological phenomena based on these models. The system may also be useful for interpreting molecular diseases and also for designing new drugs, because the disorder of cell signaling networks has close relation with molecular disease. In this paper we shall introduce the first prototype of the system.

2 System and methods

2.1 System environment

The CSNDB was implemented using ACEDB[6] system, which was originally developed by Dr. Jean Thierry-Mieg and Dr. Richard Durbin in order to manage C. elegans genome data. It is now used for wide range of other applications including various genome projects.

CSNDB was developed on workstations of Silicon Graphics Indigo. These workstations are connected to the Internet by a dedicated communication line (256Kbps).

2.2 Data sources

Knowledge on cell signaling networks was classified from a standard text book[7], reviews, and original papers. In addition we referred molecular databases on the Internet; SwissProt[8] for protein data, PDB[9] for three dimensional structures of protein, and Online Mendelian Inheritance in Man (OMIM)[10] for molecular disease.

3 Results

3.1 Representation of signal pathways

The system provides the knowledge of signal transduction pathways and of signal molecules. In CSNDB the pathways are designed to present the mechanisms of cellular signal transduction, which is the basic knowledge of the system. We considered six points regarding representation of signaling pathways;

1. diagram representation of pathway
2. hypertext link between diagram and molecular data
3. layer structure of pathway representation
4. connection between disorder of pathway and molecular disease
5. effects of external chemicals, such as toxins and drugs
6. comparison of pathways among species.

According to these points, we designed the pathway representation of CSNDB.

In CSNDB the signaling pathways are represented as flow diagrams that consist of nodes and arrows (Figure 1). A node represents a signaling molecule and an arrow represents a molecular interaction that transfers the signal. The facility of diagram depiction is provided by ACEDB, and there exist hypertext links between flow diagrams and molecular information. If one clicks a node, the next window that contains deeper information about this signal molecule is displayed, and if one clicks an arrow, then the detail modes of molecular interactions are presented as a new window. The information of molecular interaction includes following properties;

1. the mode of signal transduction; phosphorylation, ligand receptor binding, SH2 recognition, GTP/GDP exchange, and so on
2. protein sites or motifs used for signal transduction
3. character of the signal transduction; activation, suppression, perception, transduction, propagation, amplification, integration, adaptation, and response
4. cellular localization of signal transduction.

The living organisms prepare not single but plural pathways for response to stimuli. In multicellular organisms communications of pathways would be necessary to regulate complex biological mechanisms. Pathways in the cells communicate with one another in order to control biological responses as whole. Theses communications are called as cross-talk or feed-back regulations. In order to represent communications of pathways, we included the higher level pathways in CSNDB. Figure 2 shows an example of the higher level pathway. In this diagram the relation of pathways are depicted, and each node links to the individual pathways described in molecular level. In this way CSNDB contains the dynamic character of signal pathways.

Regarding as differences of pathways between species, some pathways are conserved, the others are not. Generally the higher animals possess the larger number of pathways as well as signal molecules. It seems that the diversity of pathways would be one of the driving forces of evolution from unicellular organisms to multicellular organisms. For the purpose of evolutionary analysis of pathways, we started to include information on comparisons of pathways between different species. It is still under preparation.

At present 29 pathways are included in the system.

3.2 Representation of signal molecules

Signal molecules are individual components of signal pathway. Signal molecules are grouped into eight; hormones, cytokines, receptors,

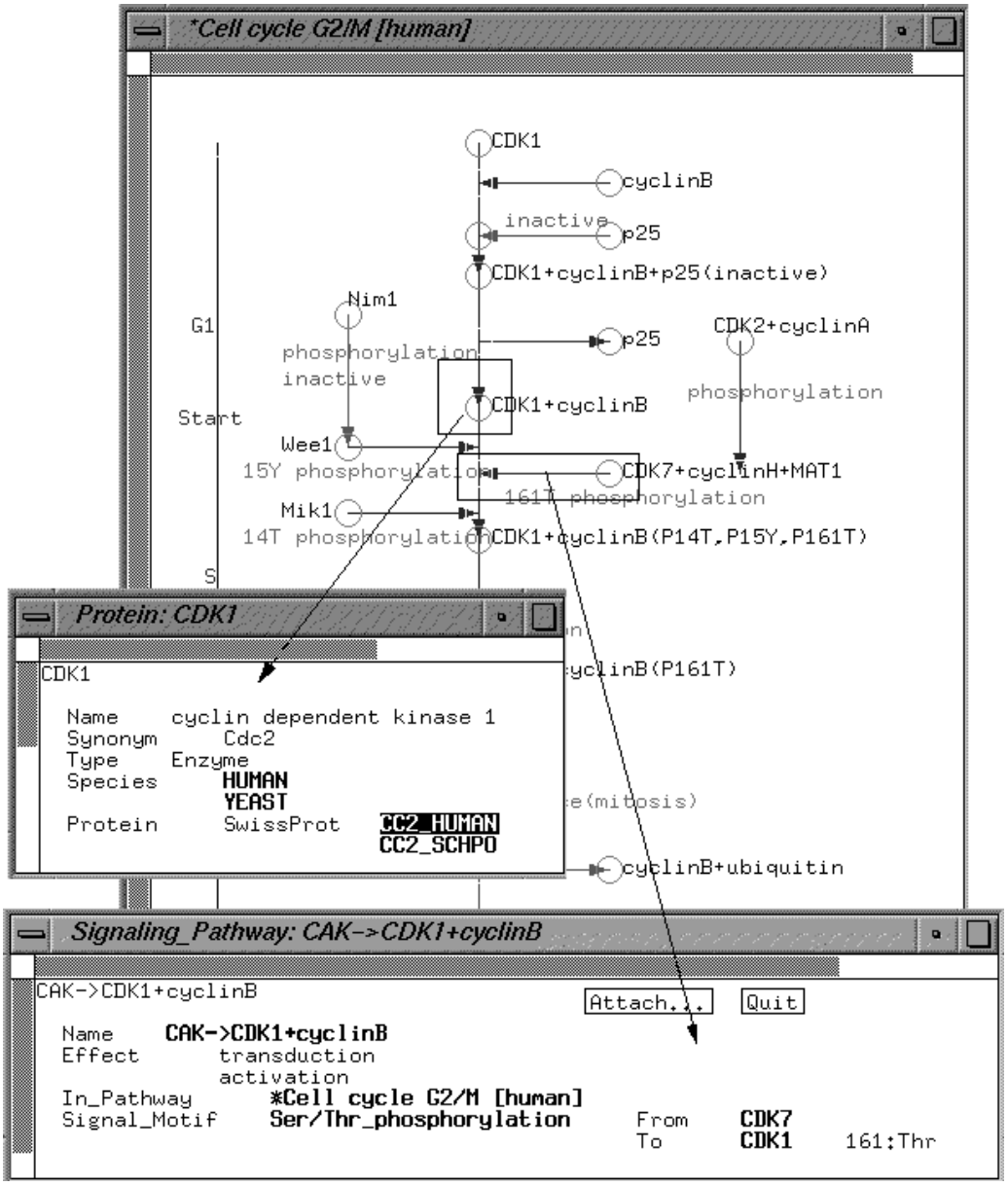

Figure 1: Representations of pathway diagrams. Left arrow indicates hypertext link between the node object and the molecular data, and right arrow indicates hypertext link between the arrow object and the modes of signal transduction.

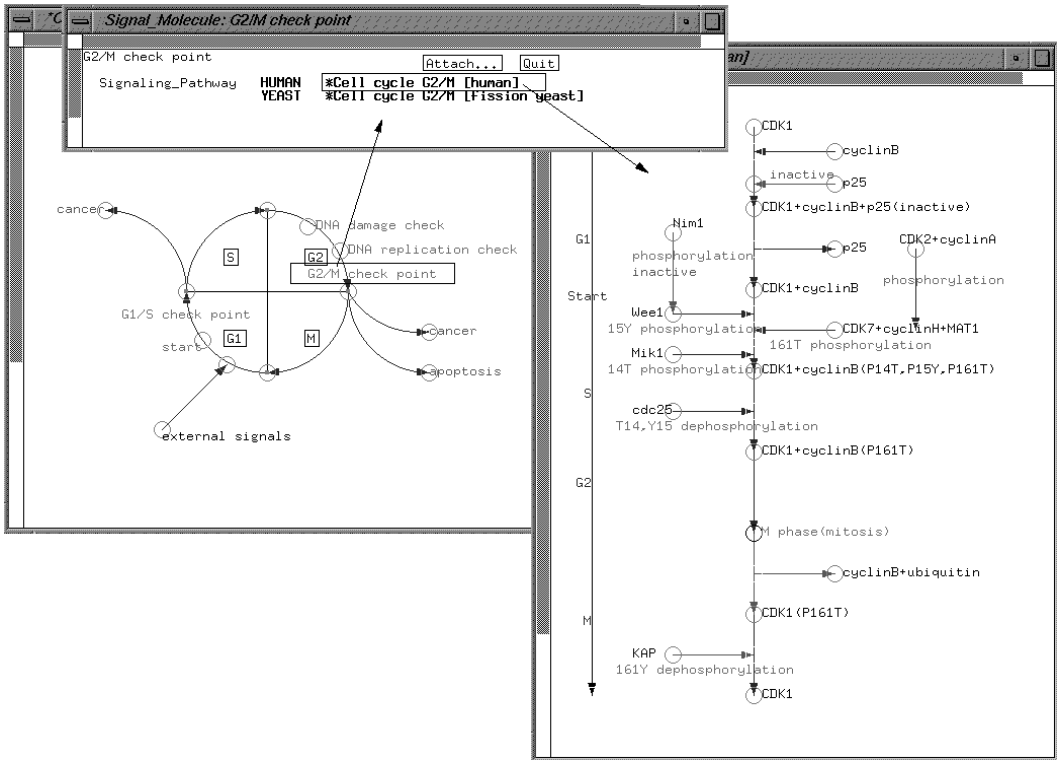

Figure 2: Representation of higher level pathways. The arrows indicate hypertext links between the node in the diagram and the component pathway described in molecular level.

ion channels, effectors, enzymes, messengers, and transcriptional factors. Each signal molecule is characterized by the following biological attributes;

1. Biochemical Data : chemical name, synonyms, cellular location, organism location, functions, isozymes, and homologues in other species.
2. Structural Data : domain structures, functional motifs, 3-D structures, and ligand bindings.
3. Genomic Data : exon and intron structures, DNA sequences, and protein sequences.
4. Associated Chemicals : activators, inhibitors, agonists, antagonists, toxins, and drugs.
5. Bibliographical Data : papers, and the date of the last update.

The ACEDB system is able to contain not only the character data but also the computer commands to invoke external programs. Using this facility we implemented appropriate external viewer programs on CSNDB. We implemented RasMol[11] for three dimensional structure, XV[12] for figure, and Netscape[13] for the connection to the Internet

Figure 3 shows examples for representations of molecular data. The chemical structure of xenobiotic ligand is displayed by XV. The three dimensional structure of protein is displayed by RasMol. This graphical model can be freely rotated and magnified. The original data provided from SwissProt is displayed by Netscape through the Internet connection.

The present CSNDB contains the following information; 281 signal molecules, 1147 sequences, 210 chemicals, 114 3-D structures, 151 pictures, and 1328 papers.

3.3 Molecular disease representation

CSNDB also contains molecular disease data. The key idea is that cell signaling networks are constructed with many oncogenes. We consider that the disorder of cell signaling networks has close relation with molecular diseases. Then we contained the information on mutational changes of signal molecules and on caused diseases.

We obtained the data of mutation sites of proteins, base changes on DNA, and molecular diseases from OMIM. In addition, we entered structural knowledge of protein. This is for the purpose of structural analysis of molecular disease. We added a functional motif or domain to each mutation site. (Figure 4). Three dimensional structures of proteins were also added (Figure 4). This structure are represented with VRML[14] model in which the mutated amino acids are particularly depicted by space fill models. Using VRML facility, we contained hypertext links

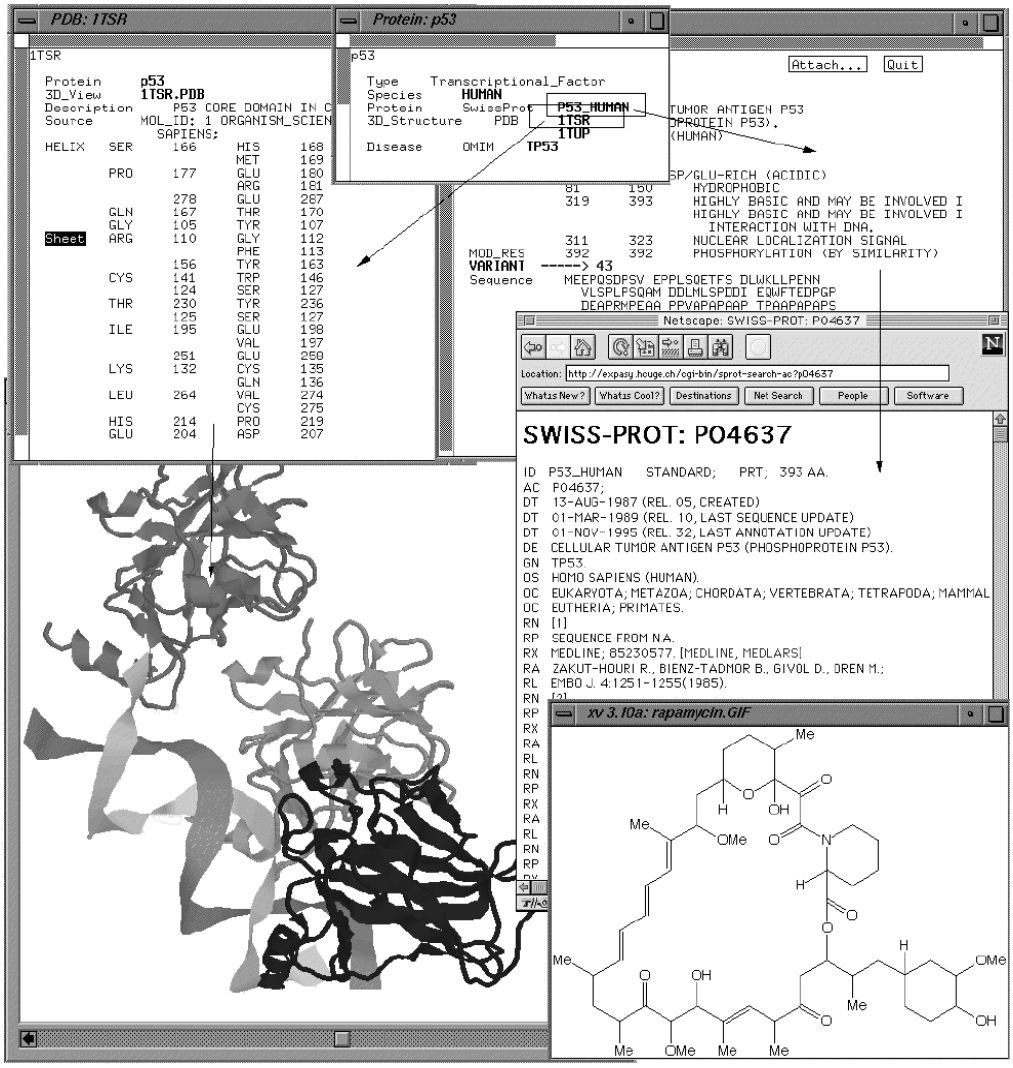

Figure 3: Presentations of molecular information of p53 (for example). Molecular data taken from original papers, SwissProt, crystal structure of p53 binding with DNA, the original data of SwissProt presented by Netscape, and chemical structure of rapamycin which inhibits progression of cell cycle within G1 phase are shown.

between individual mutated amino acids and the molecular disease data. Mutation maps were also added. They represent positions of mutated amino acid corresponding to protein sequences and functional domains (Figure 4). Since in mutation maps we utilize the same technique as pathway diagrams , there exist hypertext links between the descriptions in the maps and the disease data. Then in CSNDB, one can refer molecular disease data, three dimensional mutation sites, and mutation site maps simultaneously.

At present 18 signal molecules include molecular disease information.

3.4 Network environment

The WWW interface of CSNDB was produced using the program developed at National Agricultural Library[15]. All the facilities of CSNDB are available from the Internet (http://geo.nihs.go.jp/csndb.html).

4 Discussion

In the past decades a lot of features of biological mechanisms have been elucidated, and various biological databases have been developed. PIR[16], GenBank[17], EMBL[18], PDB, SwissProt, and OMIM are such examples. These databases have become the standard references for molecular scientists. However all these databases target the attributes of individual molecule. Novel databases for not an individual molecule but molecular interactions are required for more realistic representation of living organisms. CSNDB will be one of such examples. We consider that this kind of knowledge base for molecular interactions will be more important.

We used ACEDB for the base system of CSNDB. ACEDB is the database system originally for integrating C. elegans genome data. It is now widely used for various genome projects. ACEDB has some object oriented features, and theses features have been proved to suite for implementing various genome data. In this work we found that ACEDB was also suitable for integrating the cell signaling knowledge. Because biological data are very often incomplete and sometimes are drastically revised when new experimental facts are discovered, we consider that the flexibility of object oriented features of ACEDB are appropriate for integration of biological knowledge.

At present we included the pathways of human and yeast (fission and budding yeast). We are planing to expand the system to other species such as Drosophila and C. elegans. We are interested in the evolution of pathways, especially in the evolution from the unicellular organisms to multicellular organisms. CSNDB is expected to give good insight into this issue.

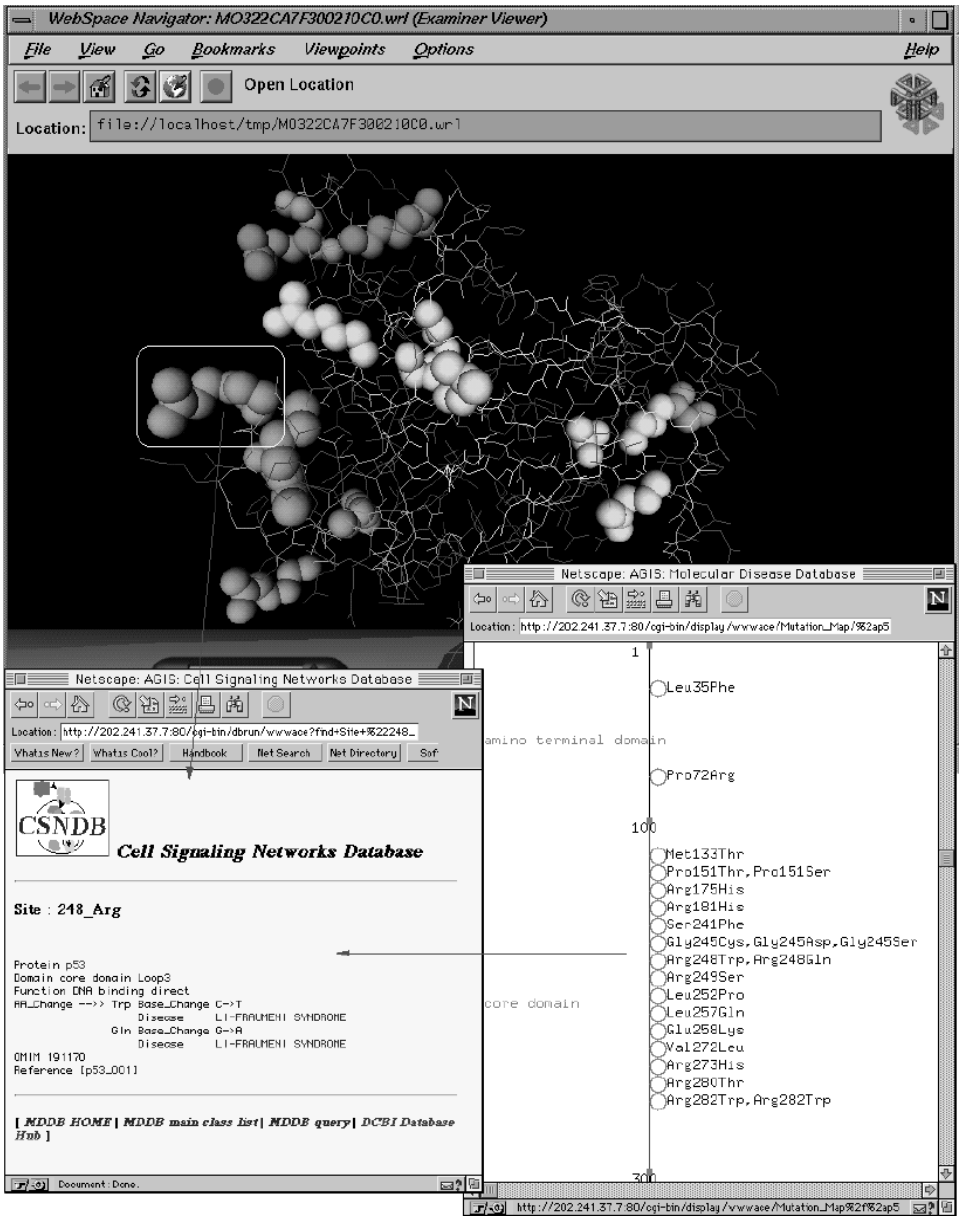

Figure 4: Representations of molecular disease data. The three dimensional image is presented using VRML environment. The arrows indicate hypertext links between the mutation site in VRML model, map presentation, and molecular disease data.

We also consider the necessity of classification of pathways in CSNDB. We need to produce hierarchical classification of entire sets of pathways in living organisms. This classification will be important for simulation of biological responses in computer system.

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

Acknowledgments

The authors thank to Dr. J. Thierry-Mieg and Dr. R. Durbin who have helped us to use ACEDB, and Dr. M. Smith who let us use OMIM data. The authors also thank to Ms. M. Yukawa and Ms. S. Hasegawa who helped for data input. This work was partly supported by Japanese Human Science Foundation Research Aid Program and by Science Research Promotion Fund from Science and Technology Agency.

5 References

1. Michal, G. Biochemical Pathways Wall Chart. Boehringer Mannheim GmbH Biochemica, 1892.
2. Karp, P. and Riley, M. and Paley, S. and Pellegrini-Toole, A., EcoCyc: Electronic Encyclopedia of E. coli Genes and Metabolism, Nucleic Acids Research, 24, 32-40, 1996.
3. Grant, D. et al. http://probe.nalusda.gov:8000/plant/aboutsoybase.html
4. Igarashi, T., Nadaoka, Y., and Kaminuma, T. A Data and Knowledge Base for Cell Signaling Networks, Proc. IPCAT95, 188-198, 1995.
5. Igarashi, T., Nadaoka, Y., and Kaminuma, T. A Data and Knowledge Base for Cell Signaling Networks, Genome Informatics Workshop 1995, 49-56, 1995.
6. Thierry-Mieg, J. and Durbin, R. http://probe.nalusda.gov:8000/acedocs/index.html.
7. Hunter, T., Cell, 50, 823-829, 1987.
8. http://expasy.hcuge.ch/sprot/sprot-top.html.
9. http://www.pdb.bnl.gov/index.html.
10. McKusick, V.A.: Mendelian Inheritance in Man. Catalogs of Human Genes and Genetic Disorders. Baltimore: Johns Hopkins University Press, 1994 (11th edition).

11. Bradley, J. ftp://ftp.cis.upenn.edu/pub/rasmol.
12. Sayle, R. ftp://colonsay.dcs.ed.ac.uk/pub/xv.
13. Netscape is a product from Communication Corporation (http://www.netscape.com/).
14. VRML is a extension of HTML which allows us virtual reality environments on WWW (http://vrml.wired.com/).
15. Barnett, J. and Bigwood, D. , http://probe.nalusda.gov:8000/acedocs/ace95//nalwww.html.
16. Sidman KE. George DG. Barker WC. Hunt LT. The protein identification resource (PIR), Nucleic Acids Res., 16, 1869-1871, 1988.
17. Benson, D., Boguski, M., Lipman, D.J., and Ostell, J. Nucleic Acids Research, 22, 3441-3444, 1994.
18. http://www.embl-heidelberg.de/.