A Preliminary Study in Data Management and Visualization of Signal Transduction Pathway Data — Viewing Interaction Data

Ken-ichiro Fukuda **Toshihisa Takagi**
ichiro@ims.u-tokyo.ac.jp takagi@ims.u-tokyo.ac.jp
Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

1 Introduction

Methodological improvements and new inventions in biological experiments, such as DNA chips, microarrays, mass spectrometry, exhaustive two-hybrid screening provides us information about the relations of genes (or its products) comprehensively. Those information are, for example, about physical or genetic interactions of proteins, clusters of genes that are co-expressed, sets of proteins of macromolecular complexes, [1, 2, 3, 4] which can be thought as components of signals in cellular signal transduction pathways. Bottom up modeling by linking up those relations will be a necessary step in systematic reconstruction of the pathways. Unfortunately, however, the actual data obtained from experiments are uncategorized and contain artifacts. Thus, an interactive data visualization tool was developed to study the nature of those interaction data.

2 System and Data

2.1 System architecture

The system was designed to visualize relations as a huge graph (Fig. 1). Each node represents a gene and each edge between nodes represents a certain relation. Users can operate the system in the following ways;

- Load multiple tables of ORFs and interaction data in CSV format.
- Calculate connected components and allows users to select components to visualize.
- Implement search of paths, edit nodes/edges and their properties.
- Save the edited graph.
- Load additional data to add nodes or edges to the edited graph.

2.2 Data Collection

The data was obtained from MIPS [5]. All ORFs with attributes such as “phenotype catalogue”,“subcellular localization catalogue”, etc. and 2458 physical interaction data (mostly obtained through two-hybrid screening and coimmunoprecipitation), 1950 genetic interaction data (mostly obtained through synthetic lethal screening) were stored in an Object-Relational DBMS (PostgreSQL [6]).

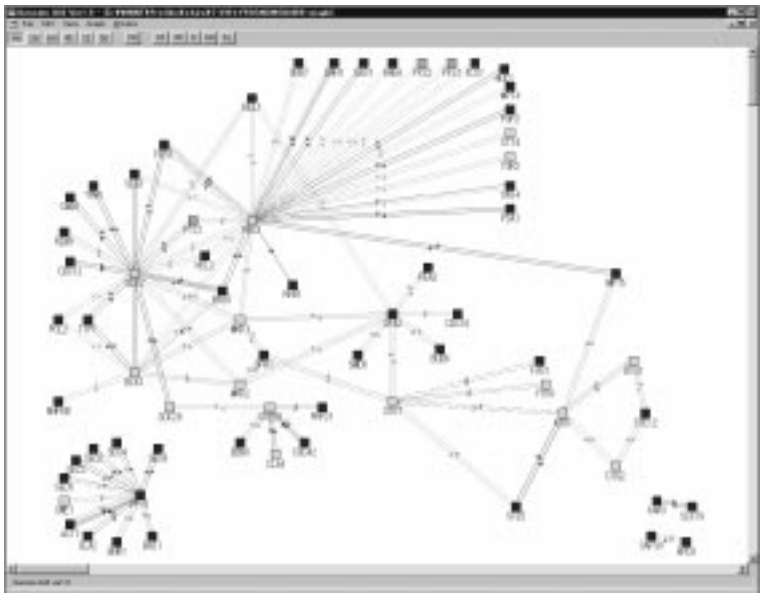

Figure 1: Example from Pkc1p-Slt2p Signaling Pathway.

3 Results and Discussion

We have designed and developed a preliminary version of an interactive data visualization tool, which represents signal transduction pathways as a graph. Detailed system architecture and experimental results will be presented at the conference site.

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.

References

[1] DeRisi, J.L., Iyer, V.R., and Brown, P.O., Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science*, 278:680–686, 1997.

[2] Doye, V. and Hurt, E.C., Genetic approaches to nuclear pore structure and function, *Trends in Genetics*, 11(6):235–241, 1995.

[3] Fromont-Racine, M., Rain, J., and Legrain, P., Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens, *Nature Genetics*, 16:277–282, 1997.

[4] Link, A.J., Eng, J., Schielts, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., Direct analysis of protein complexes using mass spectrometry *Nature Biotechnology*, 17:676–682, 1999.

[5] Mewes, H.W., Hani, J., Pfeiffer, F., and Frishman D, MIPS: a database for protein sequences and complete genomes, *Nucleic Acids Research*, 26:33–37, 1998.

[6] PostgreSQL, http://www.postgresql.org