# A Workbench for Systematic Construction of Signal Transduction Pathways: Ontologies and Architectures

**Kenichiro Fukuda** ichiro@ims.u-tokyo.ac.jp
**Mikio Yoshida** mikio@ims.u-tokyo.ac.jp
**Toshihisa Takagi** 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

To understand the phenomena of life, we need to grasp all events that occur in the life of a living cell, and then describe and comprehend the cell as a system. To that end, all proteins expressed by the cell and their functional relationships should be recorded so that researchers can build models.

Meanwhile, as a result of the recent complete sequencing of microbial genomes, DBs of complete proteomes and DBs for comparative genomics studies have been developed[1, 6, 5]. With this detailed information of all ORFs and known proteins stored in databases, researchers can now conduct systematic screening for functional analysis and acquire a large quantity of mutual interaction data[2, 4]. These data will be essential to understand the behavior of a cell as a whole system.

However conventional modeling methodology does not provide the ability to verify if the model is coherent as a biological system. Currently, researchers build their models step by step from genetic or physical binary relation data, which are mostly obtained from coimmunoprecipitation or two-hybrid screenings and synthetic lethal screenings [3].

To understand the mechanisms of a cell system, we need to verify if the stored knowledge on signal transduction pathway is correct and complement the missing links with the above mentioned data. Thus, a model construction environment with which users can conduct the following steps needs to be developed : 1. design a model interactively by accessing various kinds of resources, 2. simulate biological systems with this model, 3. debug this model or determine what it lacks, 4. verify and refine the model, 5. iterate step 1 to 4.

## 2 Ontology & Knowledge Expression

In this work we define signal transduction as the following;

- A “signaling event” is a phenomenon where an object’s change of state cause another object’s change of state[1]. Here, change of state means phosphorylation, dephosphorylation, degradation, compound formation, binding of adapter proteins (for example, proteins with an SH2 domain).
- Signal transduction is a cascade of “signaling events”, where the occurrence of one signaling event depends on other signaling events.

In this work, our cell system is represented as an event-driven system.

[1]Currently “objects” mean proteins or DNA

# 3 System architecture

The model construction system is GUI compatible and has the following features;

- pathway edit function
- truth maintenance
- links to protein DBs, transcription factor DBs and other DBs
- change view function

A "view" is a representation of a network. Users can change their view from a genetic interaction view to a physical interaction view and vise versa. Or they can switch from a signal transduction pathway view to a gene regulation network view.

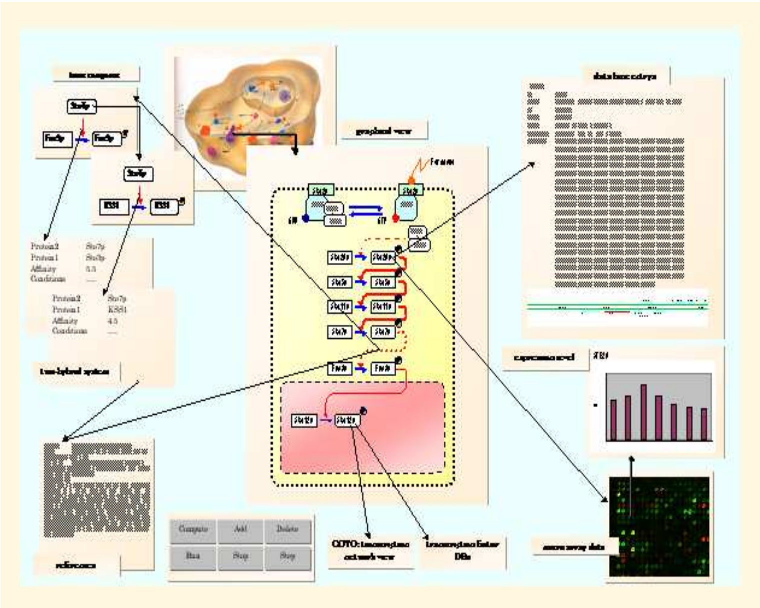

Figure 1: A snap-shot of the environment.

# References


[1] MBGD : Microbial Genome Database for Comparative Analysis, http://mbgd.genome.ad.jp/

[2] 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.

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

[4] 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.

[5] Hodges, P.E., Payne, W.E., and Garrels, J.I., Yeast Protein Database (YPD): a database for the complete proteome of Saccharomyces cerevisiae, *Nucleic Acids Research*, 26:68–72.(1998), http://www.oup.co.uk/nar.

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