OBIYagns : A GRID-BASED BIOCHEMICAL SIMULATOR WITH A PARAMETER ESTIMATOR

Shuhei Kimura[1*], Takuji Kawasaki[2], Mariko Hatakeyama[1], Takashi Naka[3], Fumikazu Konishi[1], Akihiko Konagaya[1]

[1] RIKEN Genomic Sciences Center, 1-7-22 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

[2] Fuji Research Institute Corporation, 2-3 Kanda Nishikicho, Chiyoda-ku, Tokyo 101-8443, Japan

[3] Department of Intelligent Informatics, Faculty of Information Science,
Kyushu Sangyo University, 2-3-1 Matsukadai, Higashi-ku, Fukuoka 813-8503, Japan

*To whom correspondence should be addressed.
Shuhei Kimura
Bioinformatics Group, RIKEN Genomic Sciences Center
1-7-22 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, JAPAN
Tel: +81-45-503-9302   FAX: +81-45-503-9158
e-mail: skimura@gsc.riken.jp

ABSTRACT

Summary

The OBIYagns (Yet Another Gene Network Simulator) is a biochemical system simulator that is composed of a multiple-user web-based graphical interface, ordinary differential equation solvers and parameter estimators distributed over an OBIGrid (Open BioInformatics Grid). This grid-based biochemical simulation system can achieve high performance and provide a secure simulation environment for estimating kinetic parameters in a proper time period. OBIYagns can be applied to larger system biology-oriented simulation projects.

Availability and supplementary information:

OBIYagns example models, methods, and user guide are available at: https://access.obigrid.org/yagns/

Contact: konagaya@gsc.riken.jp

INTRODUCTION

Computer simulations of mathematical models of cellular signaling networks have been recognized as useful tools for biological researchers seeking to understand the system dynamics and regulation mechanism of cell behavior and to set up subsequent working hypotheses in their studies (Hatakeyama, *et al*. 2003). Recently, several programs such as Gepasi (Mendes, 1993), E-Cell (Tomita *et al*., 1999), Virtual Cell (Loew and Schaff, 2001) and GENESIS (Sivakumaran *et al*., 2003) have been made available for such purposes. Mathematical modeling requires the appropriate values for a set of parameters (kinetic constants and physiological concentration on the cellular components), however, at present, there are no experimental methods that measure all of such parameters with accuracy. Therefore, in order to construct the model, it is essential to estimate biological parameters based on available observable wet-lab data (Swameye, *et al*. 2003).

For this purpose, biochemical simulator programs have been developed with a parameter estimation function (Mendes and Kell, 1998). However these parameter estimators do not work properly when many unknown parameters are being solved at the same time, because the parameter estimation process requires many computational resources beyond the affordable size of a single personal computer. Therefore, high performance computing systems are becoming necessary for large-scale biochemical network simulation.

The OBIYagns (Yet Another Gene Network Simulator) incorporates parameter estimators distributed over an OBIGrid (Open BioInformatics Grid) (Konagaya, 2002). The OBIYagns' web-based graphical interface lets biological researchers use the system without any knowledge of parallel processing or distributed processing, which is necessary for high performance computing based on large computational resources, due to the grid computing system (Foster, *et al*. 2001; Russel, *et al*. 1997).
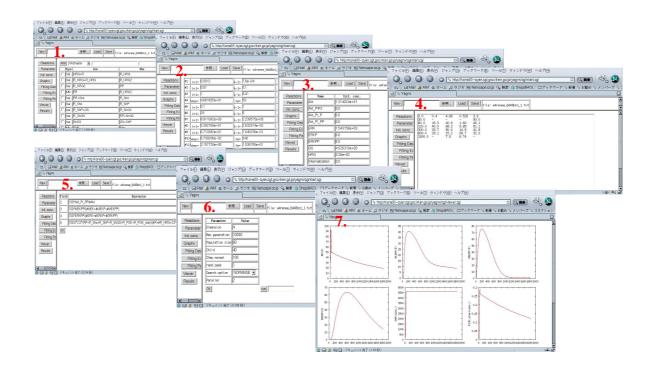
Figure1

The core of OBIYagns is composed of an ordinary differential equation (ODE) solver and a parameter estimator. It accepts as user inputs lists of reactant concentration, kinetic constants such as Mass action and Michaelis-Menten type reactions, and formulas for biochemical reactions written in chemical notation, and graphically displays the simulation results (Figure1). Biochemical reactions are automatically converted to ODEs and then solved by the ODE solver based on the numerical differentiation formula (NDF), which is suitable for solving the stiff problems that often appear in equations for biological systems (Michalski, *et al*. 1998; Stevenson, *et al*. 1984). NDF is an implicit formula that has better stability than the Gear method (Shampine and Reichelt, 1997). As for the unknown parameter estimation function, the problem for estimating unknown parameters is formulated as a function optimization problem in OBIYagns. When observed wet-data is provided, the system automatically generates an objective function of the optimization problem. The parameter estimator employs an in-house genetic algorithm called the DIDC (a genetic algorithm with Distance Independent Diversity Control) as a function optimizer. Although DIDC can solve a large number of ODEs by estimating many unknown parameters, the computational cost is high. OBIYagns is able to resolve this problem by executing DIDC in parallel in the grid (Kimura and Konagaya, 2003a and 2003b).

OBIGrid is a practical Grid system being developed by the Japan Committee on Very Large Biocomputing (VLSB, http://www.obigrid.org/vlsbio/) and the Initiative for Parallel Bioinformatics Processing (IPAB, http://www.ipab.org). The grid comprises computers connected over the Internet, that are voluntarily provided by universities, national research institutes in HPB, and enterprises in IPAB. At present, OBIGrid is composed of 492 CPUs in 27 locations. The grid is equipped with functions that are indispensable in the life sciences fields, such as database systems, computation systems, and access control systems, as well as various bioinformatics application suits protected by a VPN (Virtual Private Network). OBIYagns is expected to become a large-scale

simulation tool for biological research. Thus OBIYagns can be used for large computer simulation projects in the area of system biology.

Figure1.

OBIYagns operation screens (1) Selection of reaction types, such as mass action and Michaelis-Menten action (2) Input of kinetic parameters (3) Input of initial concentrations of reactants (4) Input of fitting data for parameter estimation (if needed) (5) Input of fitting expressions (if needed) (6) Input of GA parameters (7) Output of results as a graphical presentation

SUPPLEMENTARY DATA

Please refer to *Bioinformatics* Online.

REFERENCES

Foster, I., Kesselman, C. and Tuecke, S. (2001) The anatomy of the grid: Enabling scalable virtual organizations. *Int. J. Supercomput. Appl.*, 15, 3.

Hatakeyama, M., Kimura, S., Naka, T., Kawasaki, T., Yumoto, N., *et al*. (2003) A computational model on the modulation of MAPK and Akt pathways in heregulin induced ErbB signaling, *Biochemical J.*, **373**, 451-463.

Kimura, S. and Konagaya, A. (2003a) High dimensional function optimization using a new genetic local search suitable for parallel computers. *Proc. of IEEE Int. Conf. on Systems, Man & Cybernetics*, 335-342.

Kimura, S. and Konagaya, A. (2003b) A genetic algorithm with distance independent diversity control for high dimensional function optimization. *Trans. of the Japanese Society for Artificial Intelligence*, **18**, 193-202.

Konagaya, A. (2002) OBIGrid: Towards a new distributed platform for bioinformatics *21st IEEE Symposium on Reliable Distributed Systems (SRDS'02)*, 380-381.

Loew, L.M. and Schaff, J.C. (2001) The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol.*, **19**, 401-406.

Mendes, P. (1993) GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput. Appl. Biosci.*, **9**, 563-571.

Mendes, P. and Kell, D.B. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, **14**, 869-883.

Michalski, R., Rode, W. and Les, A. (1998) DerivFit: A program for rate equation parameter fitting using derivatives. *Comput. Biol. Med.*, **31**, 71-89.

Russel, M., Allen, G., Daues, G., Foster, I., Seidel, E., *et al*. (2002) The astrophysics

simulation collaboratory: A science portal enabling community software development. *Cluster Computing*, **5**, 297-304.

Shampine, L.F. and Reichelt, M.W. (1997) The MATLAB ODE suite, *SIAM J. on Sci.Comp.*, **18**, 1-22.

Sivakumaran, S., Hariharaputran, S., Mishra, J. and Bhalla, U.S. (2003) The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics*, **19**, 408-415.

Swameye, I., Muller, T.G., Timmer, J., Sandra, O. and Klingmuller, U. (2003) Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc. Natl. Acad. Sci. U S A.,* **100**, 1028-1033.

Stevenson, D.E., Warner, D.D. and Brown, T.R. (1984) BIOBELL – A simulation system for biochemistry and biophysics. *Comput. Biol. Med.*, **14**, 35-46.

Tomita, M., Hashimoto, K., Takahashi K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., *et al*. (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics*, **15**, 72-84.