Pacific Symposium on Biocomputing 5:314-325 (2000)

CONSTRUCTION OF A GENERALIZED SIMULATOR FOR MULTI-CELLULAR ORGANISMS AND ITS APPLICATION TO SMAD SIGNAL TRANSDUCTION

KOJI M. KYODA[†*], MICHIKO MURAKI[†‡], and HIROAKI KITANO[†§]

[†]*Kitano Symbiotic Systems Project, ERATO, JST*
6-31-15 6A, Jingu-Mae, Shibuya-ku, Tokyo 150-0001, JAPAN
{kyoda, muraki, kitano}@symbio.jst.go.jp
http://www.symbio.jst.go.jp

[*]*Department of Computer Science, Keio University*
3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-0061, JAPAN

[‡]*Medical Research Institute, Tokyo Medical and Dental University*
1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, JAPAN

[§]*Sony Computer Science Laboratories, Inc.*
3-14-13 Higashi-Gotanda, Shinagawa, Tokyo 141-0022, JAPAN

Abstract

In this paper, we report development of a generalized simulation system based on ordinary differential equations for multi-cellular organisms, and results of the analysis on a Smad signal transduction cascade. The simulator implements intra-cellular and extra-cellular molecular processes, such as protein diffusion, ligand-receptor reaction, biochemical reaction, and gene expression. It simulates the spatio-temporal patterning in various biological phenomena for the single and multi-cellular organisms. In order to demonstrate the usefulness of the simulator, we constructed a model of *Drosophila*'s Smad signal transduction, which includes protein diffusion, biochemical reaction and gene expression. The results suggest that the presence of negative feedback mechanism in the Smad pathway functions to improve the frequency response of the cascade against changes in the signaling.

1 Introduction

Recent progresses of molecular biology enable us to obtain massive data on various aspects of living systems. Although there are numbers of problems in the data generated within the current experimental methods, it is matter of time that accuracy and scopes of measurable quantities of experimental method to reach state that they can be used for biological system level study.

In order to study biological systems at the system-level, it is essential that we can use dedicated simulator that can model the essential features of

the system. To be specific, we need a generalized simulator for multi-cellular organisms which enable us to study following processes on a single system:

1. gene expression (transcription and translation)
2. biochemical kinetics (metabolism and signal transduction)
3. detailed intra-cellular model and cell-cell interaction, such as diffusion and ligand-receptor interactions

Nevertheless, we have not seen a simulator with such capability readily available to the community. There has been several previous efforts to design simulators for kinetics such as GEPASI [1], E-CELL [2]and Virtual Cell [3]. These simulators allow quantitative simulations of biochemical reactions based on ordinary differential equation. However, in order to understand higher developmental processes such as morphogenesis, it is indispensable to develop a simulator for multi-cellular organisms.

Several simulations in the Virtual Drosophila Project [4] including, eye formation [5,6], leg formation [7], wing formation [8] handle multi-cellular organisms in *Drosophila*'s developmental stages. However, they do not consider the biochemical kinetics in each cell, since they are based on Reaction-Diffusion model. The additional problem is that each simulator has been developed individually, and has been used for a specific phenomenon. There is no generalized simulator for satisfying the needs of serious biological research.This shortcoming motivated us to develop BioDrive. Table1 compares BioDrive features against other well known simulators.

The goal of our work reported in this paper is to develop a new framework for a simulator named BioDrive, which can describe biochemical reactions, gene expression, protein diffusion and cell-cell interactions. Consequently, it

Table 1: Features of the previous simulators and BioDrive. A circle symbol means that the simulator implements the property. A triangle means that the property is planned to be implemented to the simulator.

	Simulator	Kinetics	Genetics	Cell-cell Int.	Diffusion
Single Cellular	GEPASI	◯			
	E-CELL	◯	◯		
	V-CELL	◯	◯		
Multi Cellular	VDP		◯	△	◯
	BioDrive	◯	◯	◯	◯

is possible to simulate the biological phenomena, which include protein diffusion, ligand-receptor interactions, kinetic reactions, transcription/translation process at the process level, and the modified protein (e.g. phosphorylations, methylations), localization of proteins at the substance level of single/multi-cellular organisms. In order to verify this simulation system, we constructed a model of *Drosophila*'s Smad signal transduction cascade, which includes protein diffusion, biochemical reactions, and gene expressions. We believe that the Smad pathway is a good target to demonstrate the usefulness of BioDrive, because it has a feedback loop involving competition of complex formation which is typical in many cascades. The biological question is to find out the functional role of the feedback loop in this cascade.

2 System Architecture

The BioDrive consists of a simulation kernel and a graphical user interface(GUI). The kernel of the BioDrive solves the simultaneous ordinary differential equations(ODEs) at each time step using numerical methods, and computes the spatio-temporal patterns of the multi-cellular organisms and/or temporal patterns of the single cell.

2.1 Simulation Kernel

The simulation kernel of BioDrive consists of a parser and a solver. The parser reads a rule file, in which definitions of the substances and reactions of the target biological system and the simulation environment are described. Here, the term 'substance' refers to biological molecules such as proteins, mRNAs, complex of proteins, and modified proteins. The term 'reaction' represents the reaction rule between different substances. The parser also reads a initialize file, in which definitions of the distribution of the initial substance concentration are described. The texts of a rule file and initialize file are translated into the form which can be interpreted by the solver. The solver computes the biochemical reactions and gene expressions.

Parser

In order to make the solver compute protein diffusion, cell-cell interactions, biochemical reactions, and gene expressions, the parser creates lists of cells, substances, reactions, and factors. These lists are stored in the cell, substance, reaction, and factor memory, respectively. With accessing data on these memories, the solver can compute the change of substance concentration according

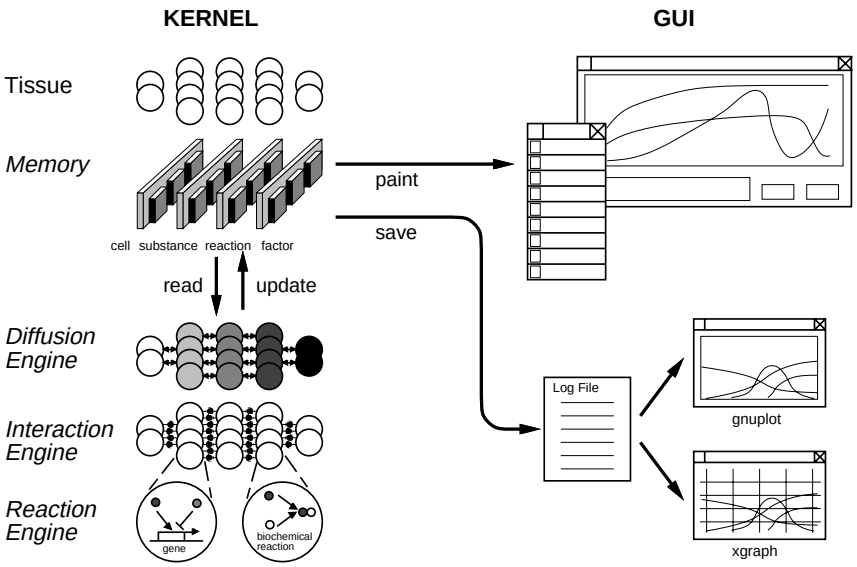


Figure 1: Architecture of the solver: The diffusion engine and the interaction engine access data on the cell memory and the substance memory. The reaction engine access data on the cell, substance, reaction, and factor memory.

to the equations of reactions. The kernel reads once the rule file into the parser, the cells which have the same concentration of substances are generated. We can change the initial concentration of the substance in each cell with the initialize file.

Solver

For simulations of multi-cellular organisms, we have built three engines: the *diffusion* engine, the *interaction* engine and the *reaction* engine. The *diffusion* engine and the *interaction* engine are designed for simulating protein diffusion and cell-cell interactions, respectively. These engines access data on the cell and substance memory. In the case of diffusion, the solver computes the concentration level of the diffused substance. In the case of cell-cell interaction, the solver accesses data on the receptor concentration and the ligand concentration in the adjacent cell. The *reaction* engine is designed for simulating biochemical reactions and gene expressions in a single cell, and also allows us to model the signaling pathways and transcription/translation processes in a

single cell. This engine access data on the cell, substance, reaction, and factor memory.

The solver is implemented as the stack machine, because the parser transforms the equations of reactions into the reverse polish notation form. The BioDrive allows us to choose the numerical methods of integration, (e.g., Euler method, fourth-order Runge-Kutta method). Figure 1 shows the solver architecture of the BioDrive.

2.2 Interface

The BioDrive GUI consists of a *Tracer Window* and a *Substance Window* implemented in C++ with X11 and XForms library. The *Tracer Window* shows the concentration of substances along time axis and/or the concentration of substances along spatial axis. The *Substance Window* defines the substances shown in the *Tracer Window*.

3 Modeling

3.1 Biochemical Reaction

In the BioDrive system, proteins, mRNAs, complex of proteins, and phosphorylated proteins are defined as substances. For example, an interaction that A makes complex with B is described as below:

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A \cdot B \tag{1}$$

where $A \cdot B$ means a complex of a protein A and a protein B. Protein A reacts with protein B to make a complex $A \cdot B$ with rate constant of k_1 (forward reaction), whereas a complex $A \cdot B$ is broken into protein A and protein B with rate constant of k_2 (reverse reaction). The velocity of producing complex $A \cdot B$ in the equation (1) is described as follows:

$$\frac{d[A \cdot B]}{dt} = k_1[A][B] - k_2[A \cdot B] \tag{2}$$

3.2 Gene Expression

Here, the gene expression means transcription and translational regulation. Since transcriptional regulation is well known, but translational regulation is not, we simplify the transcription and translation process into one black-box, which means that the amount of transcribed mRNA is the same as the amount

of resultant protein. The gene expression is represented as hill equation in this simulation. The product of gene X is described as follows:

$$\frac{d[X]}{dt} = \frac{V_{max} \cdot [S]^n}{K + [S]^n} \tag{3}$$

where n is the Hill coefficient, $[S]$ is the concentration of the activator, and where K is a constant comprising the interaction factors.

3.3 Diffusion

The secreted protein diffuses into a given concentration gradient. This long-range signaling molecule acts as a *morphogen*, specifying cell fate through membrane receptors and cellular transduction mechanisms. The diffusion equation is defined by:

$$\frac{\partial U_A}{\partial t} = D_A \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) U_A \tag{4}$$

where U_A is the concentration of secreted protein A at cell-position (x, y). D_A is a constant value which is individually determined for each protein.

3.4 Ligand-receptor Interaction

The cytoplasmic protein in a cell receives signal from ligand-receptor complexes activated by ligand, and the strength of this signal depends on the concentrations of the ligand and receptor. In this simulation system, we determine the production velocity of the ligand-receptor complex for the cell,

$$\frac{d[L \cdot R]}{dt} = \sum_{i}^{N} \left(k_1 \frac{[R]}{N} \frac{[L]_i}{N} - k_2 \frac{[L \cdot R]}{N} \right) \tag{5}$$

where $[L]$ is the ligand concentration in the adjacent cell and $[R]$ is the receptor concentration. N is the number surrounding cells, and where k_1, k_2 are reaction rate constants.

4 Smad Signal Transduction

Transforming growth factor-beta (TGF-β) is an evolutionarily conserved family of secreted factors that control cell fate by regulating cell proliferation, differentiation, matrix production, motility, adhesion, and apoptosis.

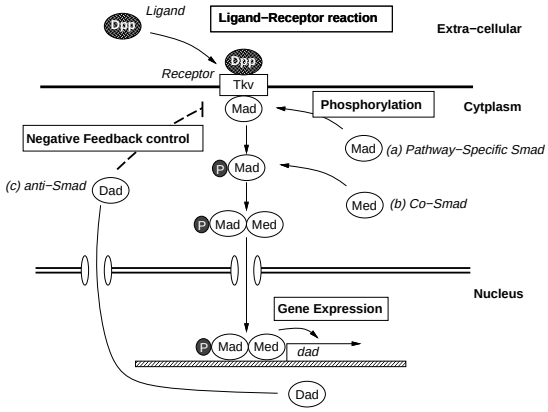


Figure 2: Smad signal transduction pathway: The signal of Dpp, which is a member of TGF-β super family, is mediated into a cell by Smad signal transduction. Smads consist of three classes of proteins; (a) receptor-regulated Smads (R-Smads), (b) the common mediator Smads (co-Smads) and (c) antagonizing Smads (anti-Smads). This signal pathway contains following biological properties: protein diffusion, Ligand-Receptor reaction, transcriptional regulation, phosphorylation, and negative feedback control.

Decapentaplegic (Dpp), a *Drosophila* member of the TGF-β family of secreted molecules, forms long-range morphogen gradients in anterior-posterior of adult appendages[9]. Dpp acts by binding to the transmembrane protein with intrinsic serin/threonine kinase activity, consisting of type II receptor Punt and to the type I receptors Thickvein (Tkv) and Saxophone (Sax)[10]. Dpp mediates signals via the Smad proteins Mother against dpp (Mad) and Medea (Med)[11]. The role of Smad proteins is to relaying signals in signal transduction by TGF-β and related factors from transmembrane receptors to the nucleus. Mad forms a heteromeric complex with Med upon phosphorylation by Tkv[11] (in this paper, the phosphorylated protein is indicated with asterisk such as Mad*, and the complex is connected with a hyphen such as Mad*-Med). Daughter against dpp (Dad) stably associates with Tkv and thereby inhibits Tkv-induced Mad phosphorylation[12].

The most interesting point of this Smad cascade is the presence of the negative feedback control (shown in Figure 2). Mad which is a co-Smad mediates extracellular signal of Dpp into the nucleus, while Dad which is an anti-Smad, competes with Mad to control the Dpp signal delicately[12]. The question is, why such a negative feedback circuit is needed. It has been suggested that the function of Dad is to stabilize the gradient of positional information emanating from Dpp-expressing cells[12]. Since the existence of such a negative

feedback circuit is a key to make clear the long-range morphogen mechanism, it is a major topic not only in the *Drosophila* development studies but in the vertebrate development studies in general.

In order to demonstrate the usefulness of BioDrive, we have implemented the Smad signal transduction pathway with BioDrive.

5 Results and Discussions

To simulate the Smad signal transduction of the *Drosophila* wing disc, we implemented the gene regulatory and biochemical networks shown in Figure 2.

5.1 Spatial and Temporal Patterning of Drosophila Wing Disc

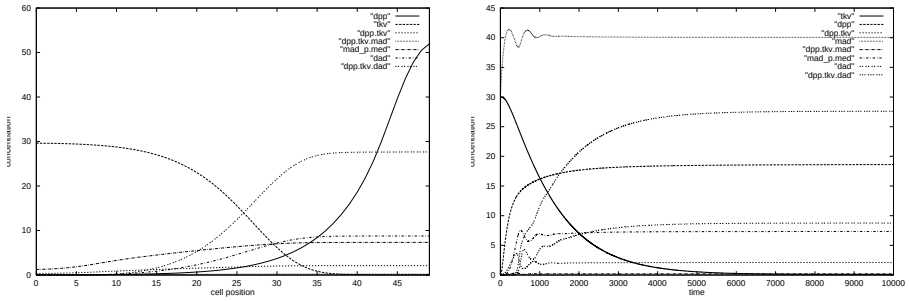
The simulation successfully reconstructed the spatial localization patterns of proteins, the complex of these proteins, and gene products (Figure 3 (a)).

In this simulation, we made *dpp* express in the anterior-posterior boundary of the wing disc as the initial condition, and Dpp diffuses to the anterior and posterior compartments. Since Dpp signaling mediated by the complex of phosphorylated Mad and Med (Mad*-Med) controls the expression of target genes (in this paper, the phosphorylated protein is indicated with asterisk such as Mad*, and the complex is connected with a hyphen such as Mad*-Med), the concentration of Mad*-Med is considered as the indicator of Dpp signal strength. In order to make cells have the positional information, Mad*-Med must be formed also in the anterior edge where Dpp signaling is weak. Figure 3 (a) shows that Mad*-Med is produced according to the Dpp signaling. The expression of *dad* which is one of the target genes of Dpp signaling, depends on the concentration of Mad*-Med, whereas it has to be highly sensitive for the concentration of Mad*-Med to form the actual Dad localization pattern. The expressions of other Dpp signaling target genes such as *sal*, *omb* should be also controled by transcriptional switching mechanism, because their expressions show all-or-none switching pattern despite the gradient of Map*-Med concentration along anterior-posterior axis is small.

We can observe not only spatio-temporal gene expression patterning but also the behavior of the single cell using BioDrive. Figure 3 (b) shows the dynamics in the single cell whose position is 40 in Figure 3 (a).

5.2 Negative Feedback Mechanism of Smad Signal Transduction

The most interesting part of this signal cascade is the presence of negative feedback mechanism. The effect of the negative feedback loop in the cascade is the inhibition of its own transcription/translation. It is generally said that



(a) spatial patterning (b) temporal patterning

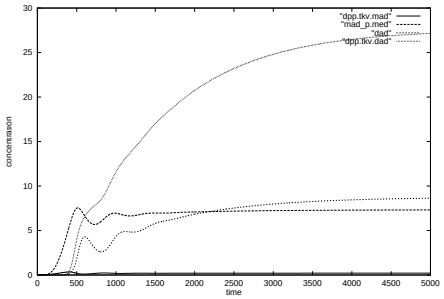
Figure 3: (a) Spatial patterning along the anterior-posterior axis of *Drosophila* wing disc. Patterning along this axis is controlled by a Dpp morphogen gradient. The left side is the anterior side and the right side is the anterior-posterior boundary. (b) Temporal patterning in the single cell.

the negative feedback mechanism of Smad cascade negatively modulate the amplitude or duration of signaling [13]. Specifically, two factors contribute to the feedback strength: (i) a time constant and (ii) inhibitor expression level. However, the systematic analysis of this feedback mechanism is difficult to verify by biological experiments. In this paper, we analyzed the Smad cascade based on the variation of the time constants using control dynamics.

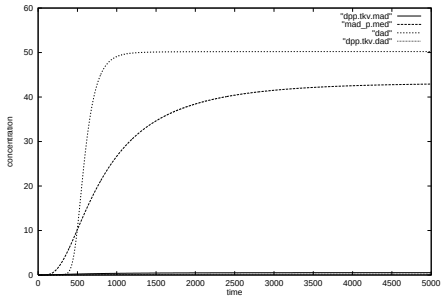
The rapid binding of Dad to Dpp-Tkv causes small time constants of the feedback system to emerge. Under this condition an initial oscillation of the Mad*-Med concentration can be observed which subsequently converges to the steady state (Figure 4 (a)). The following sequential reactions occur: (1) the accumulation of Dad represses Mad binding to Dpp-Tkv, (2) the expression level of *dad* which is the target gene of Mad*-Med becomes low, (3) the strength of negative feedback becomes weak, (4) Mad can bind to Dpp-Tkv, (5) the concentration of Mad*-Med increases, (6) Dad expresses and increases feedback level.

Conversely, the slow binding of Dad to Dpp-Tkv causes the large time constant of the feedback loop. In this situation, Mad*-Med tends to converge to the steady state without oscillatory period (Figure 4 (b)). Since the binding rates of Mad and Dad to Dpp-Tkv are well balanced, Mad and Dad do not compete to bind to Dpp-Tkv.

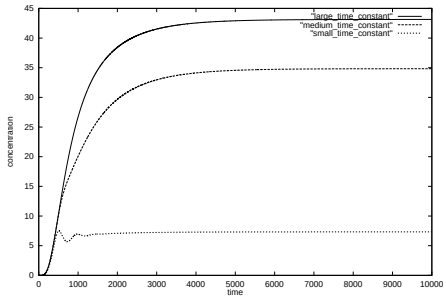
Figure 4 (c) shows the steady state levels for different time constants. Small time constants cause significantly lower Mad*-Med concentrations than large time constants. Additionally, in the system with small time constants,



(a) with the small time constant



(b) with the large time constant



(c) the responses of Mad*-Med

Figure 4: (a) The responses of the several substances with the small time constant. (b) The responses of the several substances with the small time constant. (c) The responses of Mad*-Med corresponding to the different time constants. The upper line is the response when the time constant is large. As the time constant becomes smaller, the line goes down.

the 80% threshold of the steady state of Mad*-Med is reached much faster than in a system with large time constants. Hence, the speed of complex formation of Dpp-Tkv-Dad determines the frequency response of the system which is responsible for monitoring of positional information created by Dpp.

5.3 Future Perspectives

The current version of BioDrive is able to describe all the reactions which can be represented by the ordinary differential equations. However, this requires that the number of molecules is large. BioDrive describes the gene expressions with small numbers of participating molecules by the Hill equation. Although approximate products of gene expressions can be modeled, a more detailed

mathematical analysis is desirable. The stochastic kinetic model which can consider the kinetic reactions with small number of participating molecule is shown to be more reasonable than conventional deterministic kinetics [14]. Therefore, the stochastic kinetic model is also planned to be implemented in future releases of the simulator.

Currently we hand-optimize all parameters because of the small quantity of detailed biological data. Because hand-optimization is inefficient and inaccurate, we are going to integrate an optimization algorithm to determine ideal parameter sets.

Due to space limitation, we are only able to show the frequency response which is determined by Dpp-Tkv-Dad in the Smad cascade. The expressions of inhibitors also determine the negative feedback effect, which we would like to explore in our future publications.

6 Conclusion

In this paper, we introduced the new simulation system for multi-cellular organisms, which satisfies the needs of serious biological investigations. We can model the intra-cellular and extra-cellular processes, such as protein diffusion, ligand-receptor interactions, biochemical reactions, and gene expressions. We believe that this simulation system has a capability to described all of the biological phenomena which can be modeled as the ordinary differential equations.

In order to verify this simulation system, we have implemented the *Drosophila* Smad signal transduction pathway, and reconstructed the spatio-temporal patterning of gene products, proteins, and their complexes. The results suggest that the negative feedback mechanism functions to improve the frequency response, which contributes to quickly response against Dpp morphogen positional information.

Acknowledgment

We would like to thank Dr. Shuichi Onami, Dr. Theo Sabisch, Mineo Morohashi, and Shugo Hamahashi for their advice, suggestions.

References

1. P. Mendes, "GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems", *Comput. Appl. Biosci.*, Vol.9, Num.5, 563-571, 1993.

2. M. Tomita *et al.*, "E-CELL: software environment for whole-cell simulation", *Bioinformatics*, Vol.15, Num.1, 72-84, 1999.
3. J. Schaff and L.M. Loew, "The Virtual Cell", *Proceedings of Pacific Symposium on Biocomputing '99*, 228-239, 1999.
4. H. Kitano *et al.*, "Virtual Biology Laboratories: A New Approach of Computational Biology", *Proceedings of 4th European Conference on Artificial Life*, 274-283, 1997.
5. M. Morohashi and H. Kitano, "A Method to Reconstruct Genetic Networks Applied to the Development of *Drosophila*'s Eye", *Proceedings of Artificial Life VI*, 72-80, 1998.
6. H. Ueda and H. Kitano, "A Generalized Reaction-Diffusion Simulator for Studying the Molecular Basis of Pattern Formation in Biological Systems", *Proceedings of Artificial Life VI*, 462-466, 1998.
7. K. Kyoda and H. Kitano, "Simulation of Genetic Interaction for *Drosophila* Leg Formation", *Proceedings of Pacific Symposium on Biocomputing '99*, 77-89, 1999.
8. K. Kyoda and H. Kitano, "A Model of Axis Determination for the *Drosophila* Wing Disc", *Proceedings of 5th European Conference on Artificial Life*,472-476, 1999.
9. M. Zecca *et al.*, "Sequential organizing avtivities of engrailed, hedgehog and decapentaplegic in the *Drosophila* wing", *Development*, Vol.121, 2265-2278, 1995.
10. T.J. Brummel *et al.*, "Characterization and relationship of Dpp receptors encoded by the saxophone and thick vein genes in *Drosophila*, *Cell*, Vol.78, 251-261, 1994.
11. J.D. Wisotzkey *et al..*, "Medea is a *Drosophila* Smad4 homolog that is differentially required to potentiate DPP responses", *Development*, Vol.125, 1433-1445, 1998.
12. K. Tsuneizumi *et al.*, "Daughters against dpp modulates dpp organizing activity in *Drosophila* wing development", *Nature*, Vol.389, 627-631, 1997.
13. J.L. Christian and T. Nakayama, "Can't get no SMADisfaction:Smad proteins as positive and negative regulators of TGF-β family signals", *BioEssays*, Vol.21, 382-390, 1999.
14. A. Arkin *et al.*, "Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells", *Genetics*, Vol.149, No.4, 1633-1648, 1998.