Biological System of Analysis of Systems Math

H. R. MACMILLAN, M. J. MCCONNELL, and K. TAI
Depts. of Mathematics, Pharmacology, and Chemistry and Biochemistry
University of California, San Diego
9500 Gilman Dr., M/C 0365
La Jolla, CA 92032-0365
USA
hughbob@mccammon.ucsd.edu http://mccammon.ucsd.edu/people/hmacmill/index.html

Abstract: - We present two examples of computational cellular biology that provide perspective on the complexity of creating predictive models of cell behaviors as they emerge from the interaction of molecular species. The first is a model of the diffusion and reaction of neurotransmitter in a neuromuscular junction using a continuum based finite element formulation [1]. Whereas this formulation accounts for spatial variation of neurotransmitter concentration, the single unknown, the second example is a system of ordinary differential equations to describe the biochemical state of an embryonic mouse neuron in relation to the observed behaviors of death, division, or differentiation. Each model, born of a deterministic mathematical perspective, is early in its own evolution as it grows to reflect true biology. The necessarily organic nature of these models, in simultaneously incorporating and stimulating understanding, is behind the perspective they provide on the future of computational cellular biology. Whereas in written form, this material must be presented linearly, live presentation is done in hypertext to more suitably portray the interconnected issues and methods on a multiplicity of scales.

Key-Words: - computational cellular biology, cell cycle, synapse, finite elements, dynamical systems

1 Introduction

Time offers a spectrum of scales within which complex systems are observed to emerge from the simple dynamics of interacting, and sometimes unwitting, components. A cell, as a unit of life, is near the small end of this spectrum. In the post-genomic era, annotation is underway of the specific roles gene products play within complex biochemical networks that give rise to discrete cellular behavior. This effort is creating a burgeoning interface between computational math and cell biology that presents enormous challenges if quantitative modelling is to be predictive and relevant.

To sample and subsequently survey these challenges, we present two early examples of computational cell biology: the diffusion and reaction of neurotransmitter acetylcholine (ACh) in a neuromuscular junction (NMJ) and the biochemical balance between death, division, and differentiation of embryonic mouse neurons. In exploring how the relevant biology of each example is reflected in commonly disparate branches of mathematics, the need for organic yet rigorous mathematical formalisms is exposed. In spite of being born of deterministic science, effective models shall at once evolve with and further the holistic understanding of dynamical systems important to biology. It is to this seeming paradox, the very existence of a mathematical system of analysis of systems biology, that our title lightheartedly alludes.

2 Computational Cell Biology

2.1 The Continuum Approach

Given the complexity of cellular dynamics, due to the relevance of multiple scales, nonlinear coupling, and stochastic effects, there is reluctance to believe in the prospect of predictive models of cellular behavior. Afterall, the very term “behavior” has metaphysical implications, and the applied mathematician would perhaps prefer “physiology.” For example, relative quiet, within both applied mathematics and molecular cell biology communities, surrounds a leading example of continuum based efforts, the NIH Virtual Cell Project [2]. Predating this approach is that of discrete models such as MCell [5], which generate statistics based on the probabilistic trajectories of individual molecules. Though the latter seems quite natural and is promising for small systems (such as a synapse!), it faces tremendous hurdles regarding scalability, both in terms of space and the number of distinct interacting molecular species . A compromise is the object of “stochastic” continuum based approaches, which

acknowledge "finite-numbers" effects by incorporating probablistic simulations as they quantify inherent noise due to smaller scales.

Enthusiasm for finite difference methods, as currently used in the Virtual Cell, is waning within the computational mathematics community due to the emergence of multilevel adaptive, and therefore scalable, finite element techniques. Built-in adaptive control of spatial approximation error in a variational setting yields a formulation imbued with a sense of optimality. As such, finite element formulations appear more amenable to a progressive incorporation within dynamical systems theory for coupled systems of partial differential equations — the study of the sensitivity of the solution relative to differences in prior states, especially as a function of the parameters in the equations presumed to govern. This broad claim on the wilderness of numerical analysis for solving coupled systems of partial differential equations is based solely on the flexible capacity of the variational setting to quantify relevant spatial differences between prior states, and it is merely a claim. However, regarding the stochastic nature of finitely many molecules interacting, the variational setting may again offer advantage since, by design, it allows the scientist to incorporate, directly into the approximation space used for the solution of the model, effects inferred from particle simulations such as Brownian Dynamics and Monte Carlo [5].

Although we subscribe to a finite element approach as the engine for resolving spatial effects in the first example here, we are fundamentally aligned with the Virtual Cell Project as it provides a substrate for the growth of effective continuum models. Only in the event of significant spatial complexity does the distinction between finite differences and finite elements become a relevant. In many cases, such as in our second example, spatial resolution may be prohibitively expensive to compute, impossible to verify by experiment, and/or of small consequence. Indeed there is irony in that an understanding of the extent to which each of these may be the case implicitly requires pushing the envelope of realistic models.

2.2 Synaptic Transmission

Motivation for studying the diffusion and reaction of ACh at a NMJ is threefold. Immediately, a variety of diseases are manifested at the neuromuscular junction, either in the ultrastructure of an affected NMJ or in the kinetic relationships of its molecular constituents [1]. These aspects of a model can be varied to explore, *in silica*, the effect of potentially therapeutic treatments. Of more theoretical concern, given the wealth of data on NMJs, an efficient model can be used in taxonomic coevolutionary studies of synapse ultrastructure and kinetics as each relate to muscular function. Lastly, and perhaps most importantly, synaptic transmission serves as an instructive paradigm for the maturation of mathematical models of cell biology.

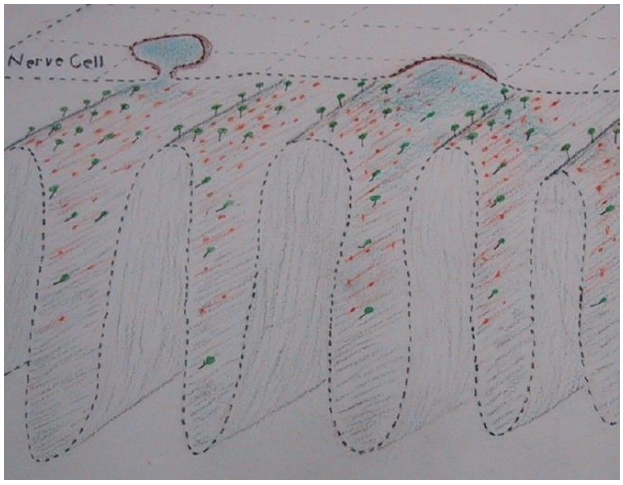

Figure 1: Artist's rendition of synaptic transmission at a neuromuscular junction.

Upon the arrival of a presynaptic actional potential at the end of a nerve cell, a cloud of ACh is released into the NMJ, pictured in Figure 1. The mechanics of this release, which involve elastic changes to the nerve cell membrane, are the first reductionist casualties in developing a *computationally* tractable model. ACh chemically diffuses across the synaptic cleft , where individual molecules encounter the enzyme acetylcholinesterase (AChe) and embedded ion channel acetylchonline receptors (AChR). When ACh binds to its receptor, the ion channel opens to permit a flux of ions into the muscle cell; this is the experimentally observable miniture endplate current (mEPC). Synchronization of this flux induces contraction of the muscle. An extremely fast enzyme, AChe provides the efficient "switch" which allows for rapidly repeating contraction of muscle.

2.2.1 Continuum Model

In an attempt to capture the physical biology of synaptic transmission, the following partial differential equation and boundary conditions for ACh concentration, $u(\mathbf{x}, t)$, has been proposed [1]:

$$\frac{du}{dt} - \nabla \cdot a \nabla u = 0 \quad \text{in } \Omega, \tag{1}$$

$$\mathbf{n} \cdot a\nabla u = 0 \quad \text{on } \Gamma/\Gamma_{AChE}, \qquad (2)$$

$$\mathbf{n} \cdot a\nabla u = -\kappa u \quad \text{on } \Gamma_{AChE}, \qquad (3)$$

with initial data $u(x,0) = u_0$, where Ω is the NMJ volume with boundary Γ, Γ_{AChE} is the disconnected surface area attributed to the enzyme, and κ is the enzyme's linear kinetic reaction constant. Note that this represents an outward flux due to consumption by the enzyme at Γ_{AChE}, and a zero flux elsewhere on Γ, including at the locations of AChR, since the receptor rereleases ACh back into NMJ. The latter condition is another part of the model subject to improvement since AChR binding of ACh may result in an "effective," albeit elusive, flux condition.

In [1], a discrete approximation to the solution of (1- 3) has been computed on several domains using a method of lines to reduce the corresponding weak form to an elliptic PDE at each time step. Given the solution at the previous time step, the elliptic PDE is solved by conjugate gradients using the finite element package FEtk [3]. In Figure 2, the finite element discretization of a realistic NMJ is reprinted from [1]. This volume mesh was generated by K. Tai based on a realistic surface representation generously provided by Bartol et al. and the MCell project [3]. This enables benchmarking the finite element solution with Monte Carlo statistics via comparison of computed miniture endplate current. However, this benchmarking is subtle because relating the approximate solution of (1 - 3) to mEPC requires accounting for the probabilistic nature of ACh binding to AChR. Currently, this is done by empirically determining a factor, α, which connects the mEPC, $I(t)$, with a surface integral of the concentration, $u(\mathbf{x},t)$, weighted by the density of AChR, $\gamma_{AChR}(\mathbf{x})$; that is,

$$\alpha \approx \frac{I(t)}{\int_{\Gamma_{AChR}} \gamma_{AChR}(\mathbf{x}) u(\mathbf{x},t) dx} \; . \qquad (4)$$

Once selected, the constant α must be shown to be valid for subsequent simulations, or else the assumption of a linear relationship, in time, between mEPC (i.e., AChR opening) and the weighted concentraion of ACh near AChR must be improved. Note that the second improvement may involve a constant time shift between the quantities. This is an aspect in which the continuum model has not reached maturation, and yet one in which its use as feedback provides insight on the actual dynamics at work. Particle methods are well suited for estimating this relationship between mEPC and $u(\mathbf{x},t)$, yet less well suited for simulations of increasing complexity. This is therefore a clear instance in which stochastic effects computed from particle methods can be incorporated into a continuum based model.

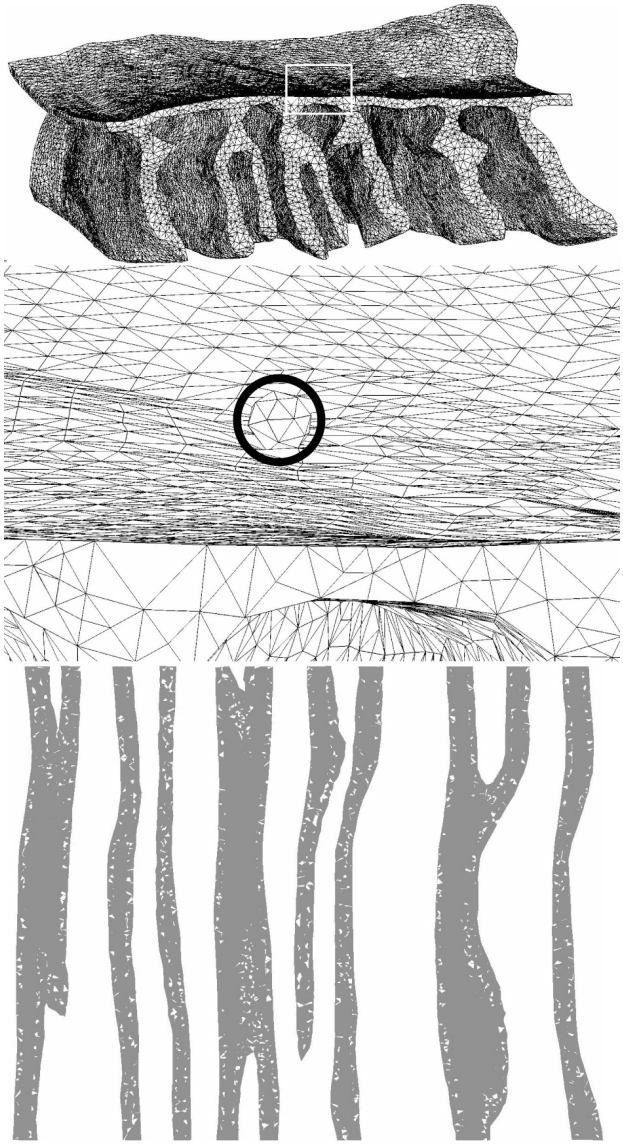

Figure 2: Tetrahedral discretization of a realistic neuromuscular junction. On top is the entire mesh, in the center is a close up of the vesicle from within the nerve cell, and on the bottom the postjunctional folds are pictured with the AChe clusters as interior boundary.

2.2.2 Future Work

Analysis is begun in [1] of the effect of NMJ architecture on muscular function, as represented in the amplitude and duration of mEPC. Specifically, differences due to muscle type (fast- or slow-twitch) as well as due to affliction with muscular dystrophy are investigated. Future studies will address physicochemical manifestations of myasthenia gravis as well as the presence of neurotoxins in NMJs. However, before

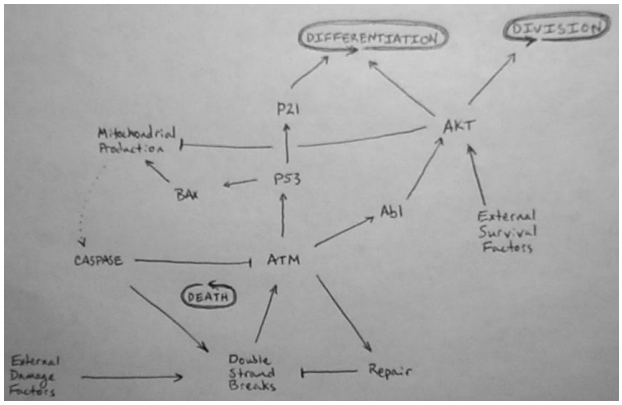

Figure 3: A reduced network of the biochemical pathways associated with the death, division, and differentiation behaviors of an embryonic neuron, as initiated by DNA damage and its activation of repair, or survival, mechanisms.

these simulations are trusted as insightful, the model must be further grounded in reality according to experimental data. This is true not only in relation to the definition in (4), but also in the representation of structural and kinetic differences. Note, however, that these are precisely the subtlties of synaptic transmission that the model is poised to explore, mirrored by evolving experimental techniques.

2.3 Neuron Development

2.3.1 Biochemical Pathways

Our second example of computational cell biology, which falls within biochemical systems theory [14], demands an increased mathematical complexity to describe coupled molecular interactions. Studies of embryonic mouse neurons suggest that a balance between DNA double-strand breaks and DNA damage signaling pathways generates genetic diversity among neurons [5-9] and influences the cellular decision to divide, differentiate, or die. In Figure 3, a network of biochemical pathways is pictured that initially captures key aspects of this dynamic. Note that this network is likely to evolve, to either expand or reduce, with insight provided by successful mathematical models and biomedical research. A given neuron with a minimal history of double-strand breaks tends to remain stationary and divide, whereas sustained levels of DNA damage are observed to trigger a cascade of events leading to cell death. However, during the transition between these states, neurons are observed to emigrate away from a proliferative zone and form the cerebral cortex, where they contribute to cognition. The hope is that a mathematical model of this system can be developed to explore the relative importance of these pathways and shed light on the structure of the network.

Motivation for the study of neuron development parallels that of our first example. First, imbalance in this dynamic is implicated as a potential root of neurodegenerative disease [8]. More theoretically, mature neurons are quite special cells in that they live throughout the organism's lifetime without becoming cancerous. Understanding the intricacies of their developmental cycle may provide insight on how to bypass blockage of cell death pathways, and thereby avoid uncontrolled and cancerous proliferation. Lastly, this example is representative of biochemical systems that will continue to confront computational mathematics in the post-genomic era.

There is a wealth of literature on various approaches to examining the behavior of biochemical networks, from graph theory to control theory to bifurcation theory, as complexity of the underlying network decreases (see [10 - 12] and references therein). Predating much of the biological interest, past studies of coupled chemical kinetics is its own mature field with a great deal to contribute to biological understanding [13]. Relative to many of these efforts, our model is premature. Its presentation below, however, provides perspective on many concerns that must be addressed for the consistent development of mathematical formulations to control the dynamics represented by biochemical pathways. Preliminary numerical results, using Matlab, will be presented in Copacabana.

2.3.2 Mathematical Model

In our previous example, complications arise from an irregular computational domain and the need to relate the solution of the model to an experimentally observable quantity. The ACh reaction with other molecular species is limited to the boundary, and hence there is no internal coupling of unknowns. Because of the inherent difficulty, alone, of a coupled system of differential equations, and because of the near impossible task of spatially resolving the governing physics, we first treat the unknowns as spatially averaged concentrations. Note that we even do this to represent the accumulation of double-strand breaks. As such, the biochemical pathways represented in Figure 3 may be interpretted mathematically as a system of coupled differential equations with time delays and general functional dependencies F and G, yet to be determined (!):

$$\frac{d}{dt}[DSB] \quad = \quad F_{DSB}([Casp],[repair],[ext_D])$$

$$
\begin{aligned}
\frac{d}{dt}[ATM] &= F_{ATM}([DSB],[Casp]) \\
\frac{d}{dt}[Casp] &= G_{mito}([BAX],[AKT],t) \\
\frac{d}{dt}[repair] &= G_{REPAIR}([ATM],t) \\
\frac{d}{dt}[Abl] &= G_{Abl}([ATM],t) \\
\frac{d}{dt}[P53] &= G_{P53}([ATM],t) \\
\frac{d}{dt}[AKT] &= F_{AKT}([Abl],[ext_S]) \\
\frac{d}{dt}[BAX] &= G_{BAX}([P53],t) \\
\frac{d}{dt}[P21] &= G_{P21}([P53],t) \qquad .
\end{aligned}
$$

The separate notation G is only to stress the relevance of a time delay due to spatial effects; this is in lieu of introducing partial spatial derivatives to arrive at a much more complicated system of coupled partial differential equations to capture the governing physics. Again, the nature of this time delay introduces further controls on the model. As before, understanding how to quantify this delay will evolve alongside targeted experiments.

We can begin to approximate the nature of coupling functions F and G using classical biochemical kinetics. According to linear theory

$$F(c_1,\ldots,c_n) = K\mathbf{c}\ ,$$

for a matrix K to be determined, and with appropriate time shifts in G. To incorporate saturation effects where necessary, Michaelis-Menten theory writes, in the case of a single argument,

$$F(c) = V\frac{c}{c+k}\ ,$$

with k some constant. Toward further sophistication in the representaion of F, a "power law" such as

$$F(c_1,\ldots c_n) = kc_1^{\alpha_1}c_2^{\alpha_2}\ldots c_n^{\alpha_n}$$

is espoused in [14]. For reactions involving multiple species, concatenated products of concentrations with differing exponents quickly lead to intractable systems of ordinary differential equations. However, the power law approach, as claimed in [14], does provide a single framework which enables the discription of many systems. In the context of approximation theory, each approach introduces parameters to represent the coupling functions that must be estimated and proven consistent with experimental understanding. However, in the spirit of constrained optimization problems, re-estimating the parameters must not compromise the proof of this consistency.

This reveals that, in essence, the goal of biochemical systems theory is to learn how to pose an inverse problem associated with a system of coupled differential equations. Since the formulation of the mathematical model is in fact the formulation of its solution, we must try to learn to pose this inverse problem well, rather than cavalierly "regularize" an "ill-posed" problem. This task is intimidating, but consolation can be found within the steady accumulation of *a priori* information regarding the biochemical states of the systems whose governing dynamics are in question. This information amounts to snapshots of the solution for all concentrations, and should of course steer the formulation of its model.

2.3.3 This Perspective

The above suggests an iterative, or cyclical, perspective in which the formulation offers a solution whose analysis informs the formulation. Instituting this perspective requires wading through a confluence of scientific disciplines to provide two crucial links: the solution of the formulation and its subsequent analysis to improve the formulation. The former entails solving differential equations numerically. The latter establishes the dual faith that the system is not overly sensitive to small variations in the parameters that define the nature of coupling, yet sensitive enough to capture significant differences in behavior.

Potential sources of inaccuracy in the formulation are multifaceted. Most fundamentally, hypersensitivity may be due to reducing a model to too few unknowns. Also, simply assuming linear, Michaelis-Menten, or power law kinetic theory may be insufficient. Another source may be associated with realizing spatial effects as a time delay instead of solving a partial differential equation explicitly (on a geometry that is itself approximate). Lastly, stochastic averaging to account for the interaction of discrete quantities is to be tuned. Therefore, understanding the relative importance of each potential source of trouble is the path to a model that is "well-posed."

Direction on this path can be obtained by sampling biochemical state space, both experimentally and computationally and in parallel, to glean their relative partitions into death, differentiation, and division. The capacity to make associations experimentally is improving through the use of high-throughput

gene expression profiling, proteomics, and even video microscopy utilizing flourescent biosensors. In turn, the ability to draw parallels to mathematical models hinges on finding tractable systems of differential equations. Conserved quantities and special qualities will be learned from experimental observations of state space, manifested according to geometry of this space, and modestly incorporated into a formulation as *a priori* information. This is a central concept in a relatively new mathematical field, the geometric integration of differential equations. For example, the qualitative behavior of many small systems of differential equations, such as nonlinear feedback loops, can be quantified as low dimensional attractors in state space.

3 Conclusion

The nonreductionist perspective of systems biology [15] is at once convincing and slippery, which is emblematic of the conundrum that life presents to science. As a movement within biology, it reflects a larger trend in science and social theory that is gaining momentum in concert with an increasing awareness of the need for sustainability on all scales of a "deep ecology" [16]: from cells to people to socioeconomic development in lieu of conflict. As such, it is hoped that an improved systems understanding of the cell can, in turn, transcend scales and offer support for, and perhaps even living proof of, the moral obligation of sustainably impacting one's own "open system" environment [17]. The role of applied mathematics (at the "burgeoning interface") is to permeate any systemic understanding of, for example, a cell, rather than reduce it and present it. Thus the goal is a biological system of analysis of systems math;, i.e., in short, biology math and not math biology.

Acknowledgments :

Hugh R. MacMillan is grateful for financial support from the Al P. Sloan/DOE postdoctoral fellowship program in computational molecular biology, and to Prof. J. Andrew McCammon, Dept. of Chemistry at UCSD, and Prof. Michael J. Holst, Dept. of Mathematics, for their postdoctoral sponsorship. Also, Drs. Steve Bond and Rich Henchman have provided fruitful discussions. Kaihsu Tai is a La Jolla Interfaces in Science graduate fellow, with partial support from the Burroughs Wellcome Fund. Mike J. McConnell is supported on a NIGMS training grant in Pharmacology from the National Institute of Health.

References:

[1] K . Tai, S. D. Bond, H. R. MacMillan, N. A. Baker, M. J. Holst, J. A. McCammon, Finite Element Simulations of Acetylcholine Diffusion in Neuromuscular Junctions, submitted to *Biophysical Journal*.

[2] L. Loew et al. , The Virtual Cell Project, *http://www.nrcam.uchc.edu* .

[3] M. J. Holst, Adaptive Numerical Treatment of Elliptic Systems on Manifolds. *Advances in Computational Mathematics* , Vol. 15, pp. 139-191.

[4] T. M. Bartol, J. R. Stiles, et al., MCell, *http://www.mcell.cnl.salk.edu/* .

[5] S. K. Rehen, M. J. McConnell, D. Kaushal, M. A. Kingsbury, A. H. Yang, J. Chun, Genetic Mosaicism in the Brain: A New Paradigm for Neuronal Diversity, *Directions in Science*, Vol. 1, pp. 53-55, 2002.

[6] D. B. Yarosh, Why is DNA Damage Signaling So Complicated? Chaos and Molecular Signaling, *Environ Mol Mutagen*, Vol. 38, pp. 132-134, 2001.

[7] P. A. Jeggo, P. Concannon, Immune Diversity and Genomic Stability: Opposite Goals but Similar Paths, *J Photochem Photobiol B*, Vol. 65, pp. 88-96, 2001.

[8] R. L. Rolig, P. J. McKinnon, Linking DNA Damage and Neurodegeneration, *Trends in Neuroscience*, Vol. 23, pp. 417-424, 2000.

[9] M. P. Mattson, Does DNA Damage Sculpt the Developing Brain?, *Trends in Neuroscience*, Vol. 24, pp. 77, 2001.

[10] C. H. Schilling, D. Letscher, B. O. Palsson, Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective, *J Theor Biol*, Vol. 203, pp. 229-248, 2000.

[11] J. J. Tyson, K. Chen, B. Novak, Network Dynamics and Cell Physiology, *Nat Rev Mol Cell Biol* , Vol. 2, No. 12, pp. 908-916, 2001.

[12] H. De Jong, Modeling and Simulation of Genetic Regulatory Systems: A Literature Review, *J Comp Biol*, Vol. 9, No. 1, pp.67-103, 2002.

[13] R. J. Kee, F. M. Rupley, J. A. Miller, CHEMKIN software, Sandia National Laboratories, *http://www.ca.sandia.gov/chemkin/* .

[14] E. O. Voit, *Computational Analysis of Biochemical Systems*, Cambridge University Press, 2000.

[15] L. Hood, et al. , *http://www.systemsbiology.org* .

[16] F. Capra, *The Web of Life*, Anchor Books, 1996.

[17] L. V. Bertalanffy, *General Systems Theory: Foundations, Development, Applications*, George Braziller, Inc. , New York, 1968