The *Karyote*® Physico-Chemical Genomic, Proteomic, Metabolic Cell Modeling System

**P. ORTOLEVA, E. BERRY, Y. BRUN, J. FAN, M. FONTUS, K. HUBBARD,
K. JAQAMAN, L. JARYMOWYCZ, A. NAVID, A. SAYYED-AHMAD, Z. SHREIF,
F. STANLEY, K. TUNCAY, E. WEITZKE, and L.-C. WU**

ABSTRACT

Modeling approaches to the dynamics of a living cell are presented that are strongly based on its underlying physical and chemical processes and its hierarchical spatio-temporal organization. Through the inclusion of a broad spectrum of processes and a rigorous analysis of the multiple scale nature of cellular dynamics, we are attempting to advance cell modeling and its applications. The presentation focuses on our cell modeling system, which integrates data archiving and quantitative physico-chemical modeling and information theory to provide a seamless approach to the modeling/data analysis endeavor. Thereby the rapidly growing mess of genomic, proteomic, metabolic, and cell physiological data can be automatically used to develop and calibrate a predictive cell model. The discussion focuses on the *Karyote*® cell modeling system and an introduction to the *CellX*® and *VirusX*® models. The *Karyote* software system integrates three elements: (1) a model-building and data archiving module that allows one to define a cell type to be modeled through its reaction network, structure, and transport processes as well as to choose the surrounding medium and other parameters of the phenomenon to be modeled; (2) a genomic, proteomic, metabolic cell simulator that solves the equations of metabolic reaction, transcription/translation polymerization and the exchange of molecules between parts of the cell and with the surrounding medium; and (3) an information theory module (ITM) that automates model calibration and development, and integrates a variety of data types with the cell dynamic computations. In *Karyote*, reactions may be fast (equilibrated) or slow (finite rate), and the special effects of enzymes and other minority species yielding steady-state cycles of arbitrary complexities are accounted for. These features of the dynamics are handled via rigorous multiple scale analysis. A user interface allows for an automated generation and solution of the equations of multiple timescale, compartmented dynamics. *Karyote* is based on a fixed intracellular structure. However, cell response to changes in the host medium, damage, development or transformation to abnormality can involve dramatic changes in intracellular structure. As this changes the nature of the cellular dynamics, a new model, *CellX*, is being developed based on the spatial distribution of concentration and other variables. This allows *CellX* to capture the self-organizing character of cellular behavior. The self-assembly of organelles,

Center for Cell and Virus Theory, Indiana University, Bloomington, Indiana.

viruses, and other subcellular bodies is being addressed in a second new model, *VirusX*, that integrates molecular mechanics and continuum theory. *VirusX* is designed to study the influence of a host medium on viral self-assembly, structural stability, infection of a single cell, and transmission of disease.

INTRODUCTION

THE COMPLEXITY OF THE CELL has led to the development of models that focus on a subset of the operating processes. Examples include glycolytic models and maps of relationships between specific genes and enzymes. However, the availability of gene sequence, metabolic network, structural, and other data gives credence to the notion of comprehensive cell modeling. In this manuscript, we briefly review the history of cell modeling, present progress attained in the last few years at the Center for Cell and Virus Theory and conclude with new directions in the development of mesoscopic cell and virus models that bridge the molecular-to-whole-cell scales of processes in an attempt to develop a predictive and comprehensive cell model. Our most well-developed model is *Karyote*®; through a web-based implementation, it integrates cell modeling with a variety of data types to arrive at an automated calibration and risk assessment methodology. In our next generation cell model, also part of these larger efforts (denoted *CellX*®), we are accounting for an even wider range of physical and chemical processes than those included in the *Karyote*® package. Both *Karyote*® and *CellX*® can be run as stand-alone cell modeling packages. They also will have interfaces with, and can be disassembled for use in, other efforts or for directly accepting their experimental data formats.

Finally, we briefly present the *VirusX*® simulator. This effort, independent of *Karyote*® and *CellX*®, is being undertaken to explore the mesoscopic nature of cellular and sub-cellular dynamics—that is, the interplay between atomic and overall cell scale behaviors. Our goal is to provide a selection of models that address the various length and timescales relevant for predicting the behavior of cells. Applications include the discovery of drugs and treatments, methods for monitoring design of microbes for biotechnical functions, stem cell research, and predicting emergent bacterial and viral strains for security and global health considerations.

Activity in quantitative cell modeling dates back over half a century. Rashevsky (1960) described simple cell models with reaction and transport. Turing (1952) showed that compartmentalized models could display self-organizational behavior (see also Prigogine and Lefever, 1968). A two-box model of cell division was used to show how asymmetric differentiation could occur upon division (Ortoleva and Ross, 1973a,b). It was shown that irreversible transitions in the state of a cellular reaction-transport system could occur in association with disconnected branches of multiple steady states of a metabolic network (Hahn et al., 1973) (Fig. 1). Metabolic control analysis was introduced at this time (Kascer and Burns, 1973) and many metabolic simulators for steady-state conditions were developed based on this approach (e.g., Gepasi [Mendes 1997] and SCAMP [Sauro 1993]). Models were developed to account for intracellular diffusional gradients; early results illustrated calcium waves and electrophysiological self-organization (Larter and Ortoleva, 1981, 1982; Ortoleva 1981a,b, 1992).

Interest in the stochastic nature of reaction-transport systems dates back to McQuarrie (1976), who set forth a master equation for the analysis of fluctuations in reacting systems. Early studies showed how fluctuations could significantly modify rate laws (Brennig et al., 1976). Fluctuations were shown to affect the nonlinear dynamics of small systems using a molecular dynamics approach (Ortoleva and Yip, 1976; Delle Donne and Ortoleva, 1978). A cellular metabolic network that supports multiple steady states in the macroscopic (i.e., fluctuation-free) equations could experience limit cycle behavior in the fluctuation-renormalized system (Hahn et al., 1974). A variety of authors have investigated the influence of fluctuations in reaction-transport systems (Gillespie, 1976, 1977; McAdams and Arkin, 1997; Firth and Bray, 2000; LeNovere and Shimizu 2001; Shimizu and Bray, 2001); also, the role of spatial dimensionality and deviations from local mean field theory due to fluctuations was delineated (Delle Donne and Ortoleva, 1978).

Accounting for the complexity of the cellular metabolic network was pioneered in work on glycolosis (Chance et al., 1964; Chance et al., 1973) and early development at the single-cell stage (Larter and Or-
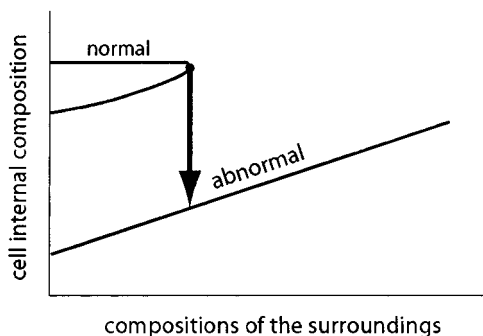
FIG. 1. The existence of isolated branches of steady-states in the cellular reaction-transport network introduce the possibility of irreversible cell transformation (e.g., the normal cell state could be on the isolated branch).

toleva, 1981, 1982). More recent investigations focused on the detail at which the metabolic network was described (Goldbeter, 1996).

Integration of metabolic models with genomic and proteomic ones has resulted in the beginnings of comprehensive cell simulators: E-Cell (Tomita et al., 2001); Virtual Cell (Schaff et al 2001); JigCell (Tyson, 2001); and Karyote® (Weitzke and Ortoleva, 2003; Navid and Ortoleva, 2003; Sayyed-Ahmad et al., 2003). The *Karyote* model and its extensions are the focus of this article; it is the only model wherein transcription and translation are described by a polymerization chemical kinetic formulation so as to naturally integrate the genome and proteome with the metabolic chemical kinetics (Fig. 2). Other unique aspects of *Karyote* are the rigorous handling of reactions that occur on a wide range of timescales to arrive at a very general equilibrium and steady-state effective rate laws. In the latter case, our multi-scale formulation accounts for
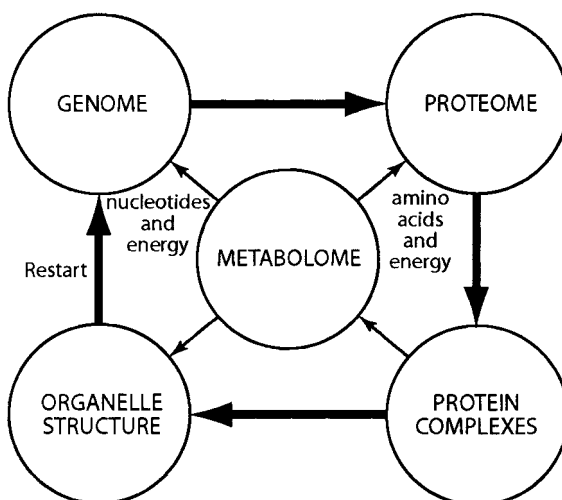

FIG. 2. The family of cell models being developed at the Center for Cell and Virus Theory accounts for the interplay of a complete set of subsystems. These include the following: genome (replication and transcription through templated polymerization kinetics); proteome (translation through templated polymerization kinetics and post-translational processes to make proteins); metabolome (reaction-transport biochemical kinetic theory); protein complexing to form enzymes, ribosomes, fibrils and other meso-structures through mixed all-atom and field theory approaches); and organelle structure (self-assembly and division of nuclei, mitochondria, lysosomes, proteomes, etc. through mesoscopic field evolution equations).
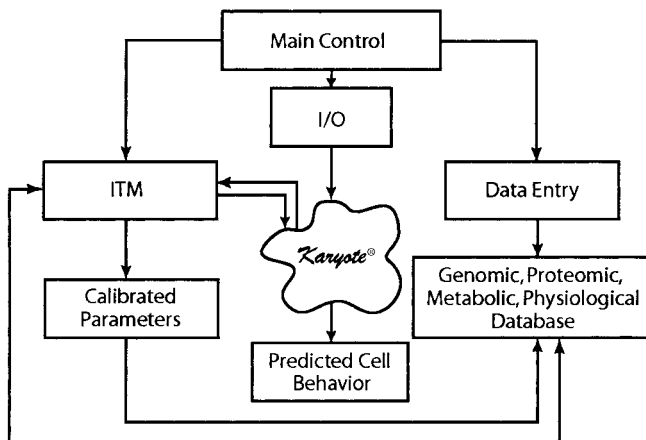
FIG. 3. Flowchart for our web-based cell model. The model can be run remotely (http://biodynamics.indiana.edu).

the conservation of mass of low concentration molecules that mediate enzyme activity or gene control. As the *Karyote* model is fully kinetic, it can capture cellular oscillatory dynamics, phenomena not captured in MCA or MFA simulators.

In the sections to follow, we present more details on the *Karyote* modeling system as depicted in Figure 3. Our self-organizing cell model, *CellX*, and simulator for viruses and subcellular bodies, *VirusX*, are also presented.

DISCUSSION

The Karyote® compartmentalized physico-chemical model

The physico-chemical processes on which the *Karyote* cell simulator is based are summarized in Figure 2. We have tested our model for glycolosis in *Trypanosoma brucei*. Our results are in better agreement with experimental results than previous studies (Navid and Ortoleva, 2003). Since the cell simulator is fully kinetic, as opposed to a Metabolic Control Analysis formulation, *Karyote* can capture nonlinear dynamics phenomena such as the complex metabolic oscillations of Figure 4.

Karyote accounts for an extensive set of processes and incorporates many features:

- General finite rate and fast (equilibrated) reactions
- Minority species (e.g., enzymes) with the associated steady-state cycles (Fig. 5) of arbitrary complexity (including cofactors and side branches)
- Compartmentalization whereby processes take place in appropriate zones (e.g., cytoplasm or organelles) and exchange of molecules between them
- Genomic and proteomic modules wherein gene sequence data is used to generate transcription and translation via templated polymerization kinetics, which, for prokaryotes, accounts for the coupling of transcription and translation (Fig. 6)
- Control and regulation of gene expression (Fig. 7)
- Equations for computing the electrical potential in each compartment and passive as well as active transport between them
- Membrane-localized processes (e.g., ion pumps or membrane-bound enzyme reactions) that involve molecules on both sides of the membrane and within it
- Multiple timescale techniques
- Web-based graphical I/O (http://biodynamics.indiana.edu)
- Integration into the cell modeling system (Fig. 3)

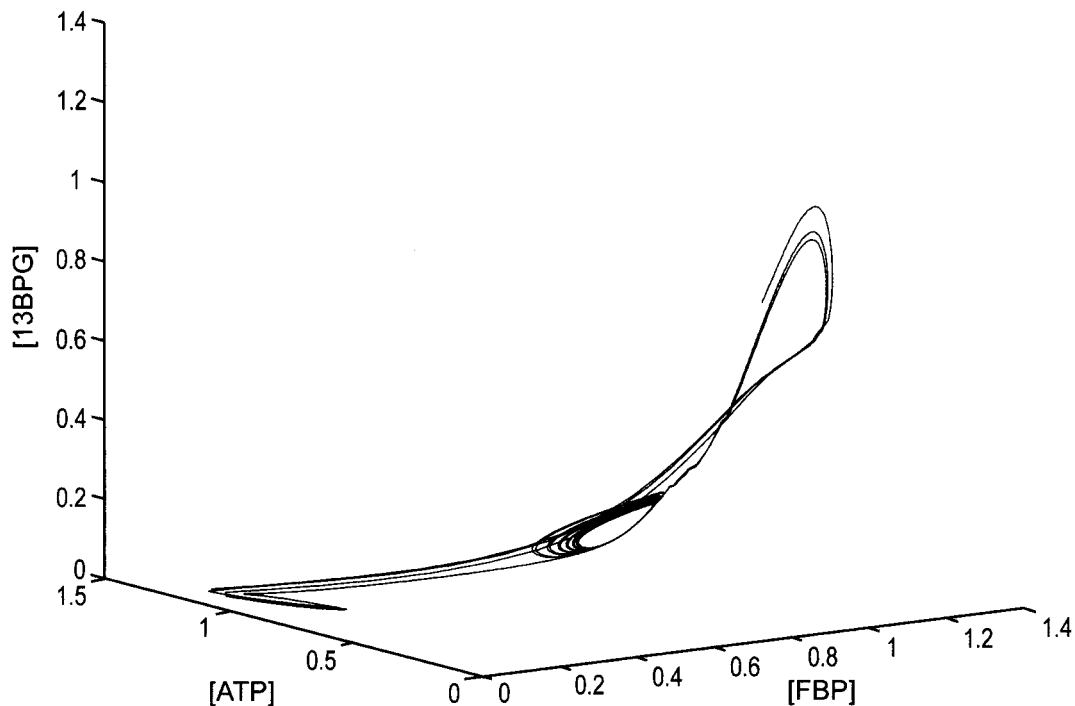As the *Karyote* system is a work in progress, further features are continuously being added.

FIG. 4. Oscillatory dynamics in a 59 species metabolic network including 28 fast and 11 slow reactions for a muta-tion of *T. brucei*. The graph illustrates the associated orbit in the space 13BPG, ATP and FBP concentrations (in mM). (Adapted from Navid and Ortoleva, 2003.)

Karyote accounts for the multiple timescale nature of cell dynamics. The multiple timescales have sev-eral origins. The most obvious is the existence of reactions that have large rate coefficients so that they are maintained close to equilibrium (e.g., $H^+ + OH^- \Leftrightarrow$ water). Cycles of fast reactions may be in steady state balance as suggested in Figure 5; for such reaction networks, generalized steady-state rate laws are con-structed automatically via a rigorous multiple-scale analysis.

The multiple timescale analysis built into *Karyote* proceeds as follows. Let ε be the ratio of the short to long timescales or the typical minority to majority species concentrations. In a series of studies (e.g., Or-toleva and Ross, 1973; Ortoleva, 1992; Weitzke and Ortoleva, 2003; Fan et al., 2003), we examine the be-
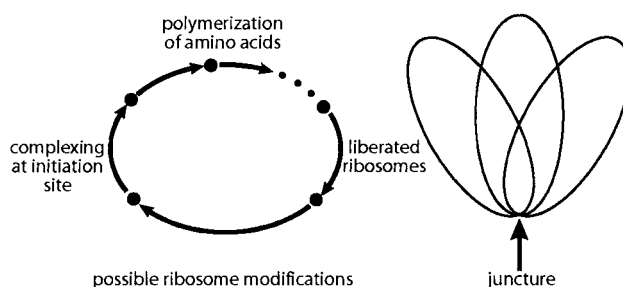

FIG. 5. (**a**) In a complex biochemical network, an enzyme or other factor (here a ribosome) can undergo a many-step cyclic sequence of transformations. (**b**) In *Karyote*®, the kinetic cycle can have a multi-lobed or more complex structure that leads to generalized steady-state kinetics automatically determined from the structure of the stoichio-metric matrix associated with fast equilibrium/steady-state reactions. (Adapted from Weitzke and Ortoleva, 2003.)
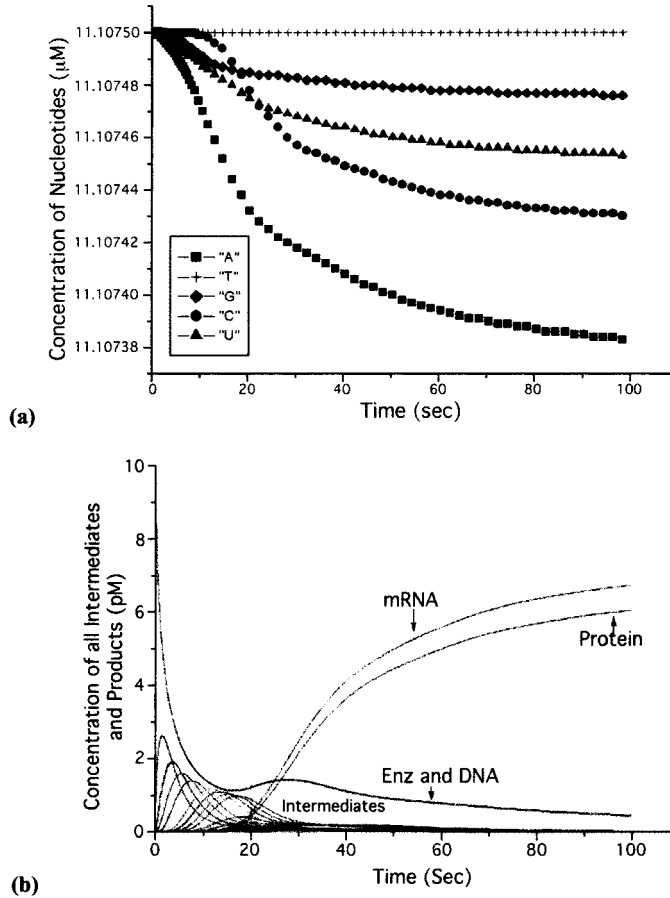
(a)

(b)

FIG. 6. *Karyote*® predicted dependence of nucleotide concentration on time during transcription for the gene TTCT-TATACGGGACT. As nucleotides are depleted, *mRNA* synthesis slows down, illustrating an important feature of coupled cellular dynamics captured in *Karyote*®. (**b**) Evolution of concentration of *DNA*, *RNA* Polymerase II and complexes involved in transcription/translation as well as *mRNA* and protein (before post-translational modification) created under the same condition as seen in a. (Adapted from Weitzke and Ortoleva, 2003.)

havior of reaction-transport systems in the asymptotic limit $\varepsilon \to 0$. A set of equations is obtained that can be efficiently solved numerically, while the full equations for finite but small ε cannot be solved with direct numerical approaches for practical execution times.

Transcription and translation are modeled using a polymerization kinetics approach. Simple polymerization of amino acids takes place by monomer addition (i.e., $(AA)_n + AA \to (AA)_{n+1}$). However, in translation, this process is guided by *mRNA* catalyzed by ribosomes, facilitated by *tRNA*, and controlled by cofactors. This complexity is accounted for in *Karyote*® (Weitzke and Ortoleva, 2003). As translation and transcription involve the consumption of nucleotides and amino acids, these processes are strongly coupled to the metabolome in *Karyote* (Figs. 1, 6, and 7).

Model-building and data archiving

Data on the wide variety of processes needed to run a *Karyote* simulation is entered through a web-based interface (Fig. 3). Input parameters are of two types (i.e., those for the reaction-transport laws and those needed to specify the structure of the cell). Input data includes the following:

- Metabolic reaction rate and equilibrium constants (fast or slow and with or without minority control species)
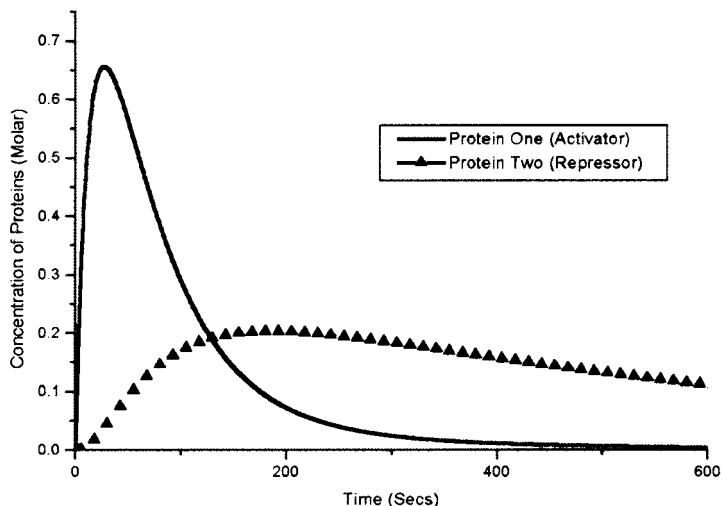
FIG. 7. Time-dependence of the concentration of Protein One and Two during gene controlled coupled transcription and translation. Gene One produces Protein One, which, in turn, activates Gene Two; Protein Two represses Gene One, ultimately resulting in a steady-state of the coupled genome/proteome dynamics in this two-gene system. (Adapted from Weitzke and Ortoleva, 2003.)

- Rate and equilibrium constants that may be different for each compartment
- Compartment structure (volume, surface area, connectivity to other compartments)
- Membrane transport properties
- Chemical species charge
- A method for archiving the data for the specific cell type of interest accomplished by entering cell type-specific data and graphically manipulated universal reactions
- Preserving the integrity of an individual user while allowing others the ability to share data in a dynamically growing database of the above cell properties

The last feature provides a platform for networking a community of biological, medical and biotechnical researchers.

To make the data archive dynamically integrated with the other *Karyote* modules, information can be automatically harvested and formatted as a cell simulator input file. For example, in running the cell simulator one may choose a specific cell type and the system automatically gathers all data from the *Karyote*® archive needed to run a simulation. Also, new rate or other parameter values obtained from calibration using information theory may be automatically added to the database along with information on the individual who carried out the calibration and the experimental data used for it (Fig. 3).

Information theory module (ITM): automated calibration, uncertainty analysis, and addressing model incompleteness

Considering the complexity of the network of reactions and of intracellular structure, as well as uncertainty in the rate laws for the participating reaction and transport processes, it seems evident that a model of cell behavior must, in some way, be cast in terms of probability theory. Furthermore, the small numbers of macromolecules or their complexes (e.g., enzymes and ribosomes) implies that key aspects of cellular modeling must be carried out within a statistical framework. Thus, the *Karyote* cell modeling system includes an information theory component for constructing the probability of cell state and model parameters for calibration, model development and risk/uncertainty assessment. In the information theory formulation, we construct the probability of cell reaction-transport parameters and the time course of given user-specified concentrations. The maximum of this probability gives the most likely value of these quantities while the overall shape of the probability can be used to assess the uncertainty of the predictions of descriptive variables and phenomenological parameters.

The starting point for the ITM is the entropy S set forth as a measure of overall uncertainty (Shannon and Weaver, 1949; Jaynes 1957; Ortoleva, 2003):

$$S = -k\mathbf{S}\rho\ln\rho \tag{1}$$

where k is a positive constant, $\rho(\Gamma)$ is the probability, Γ is a set of variables that are considered uncertain, and $\mathbf{S}$ denotes functional integration. In the *Karyote* system, these include the following:

- Reaction rate and equilibrium constants
- Membrane reaction/transport properties
- Gene control, translation/transcription parameters
- The time-dependence of chemical species for which the mechanism of creation and destruction are yet unknown

Due to the last factor, ρ is seen to be a function of the time course of these concentrations as they change over the cell cycle or in response to modifications in the surrounding medium (Sayyed-Ahmad et al., 2003).

The information theory prescription is to admit the greatest uncertainty (i.e., maximum S) constrained only by the available information. For cells, these include those suggested in Figure 8. Constraints on the maximization of S are of several types:

- Measures of error between model-predicted and observed NMR, spectroscopy, and other data
- General information we have on the range of rate coefficient and other physico-chemical parameters
- Regularization conditions on the rate of change of chemical species concentrations
- Statistical mechanical constraints on the magnitude of thermal fluctuations

Let $C_1, C_2, \ldots C_{N_C}$ be the set of N_C constraints, which by construction have zero average. Then ρ is the functional that maximizes

$$S - \sum_{l=1}^{N_C} \beta_l C_l \tag{2}$$

for Lagrange parameters $\beta_1, \beta_2, \ldots, \beta_{N_C}$. The latter are determined from the constraints (Sayyed-Ahmad et al., 2003). Having constructed $\rho(\Gamma)$ we then seek the value of Γ that maximizes ρ. The associated uncertainty Γ in is also computed in the ITM. For example, let $\Psi(t)$ be the set of unknown concentration time courses whose dynamics are not accounted for in the model (i.e., in the reaction network). The most probable $\Psi(t)$ is that which maximizes ρ:
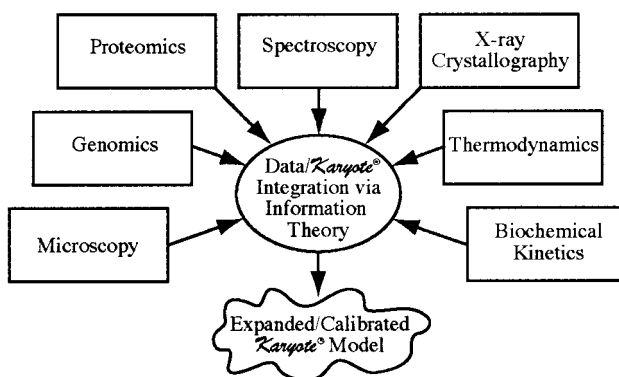
$$\frac{\delta\rho}{\delta\Psi(t)} = 0 \tag{3}$$

FIG. 8. Multiple types of data are integrated via our information theory methodology to automatically yield improvements as new data becomes available. This is already implemented in *Karyote*® and will also be extended to *CellX*® and *VirusX*®.

This functional differential equation is solved numerically as is the uncertainty envelope via a functional Gaussian approximation (Sayyed-Ahmad et al., 2003). Technical difficulties arising from the paucity of data and conflicting data are surmounted by building expertise into the constraints. Ideal experimental data to be built into the constraints are time series data on the proteome or metabolome.

The CellX® simulator

Improvement of cell modeling beyond that of the spatially uniform, fixed compartmented approach already implemented in *Karyote* requires several advances:

- Allowing the compartmental structure to be dynamic—that is, membranes can emerge, move or disappear via laws of structural self-assembly
- The description must be augmented to account for greater detail in molecular structure and orientation
- Composition within each compartment must be allowed to be non-uniform and to evolve via equations of mass conservation

As an example of the type of phenomena that are to be accounted for in our next generation cell model, the life cycle of *Caulobacter crescentus* is shown in Figure 9. *Caulobacter* cells undergo morphological changes that allow visual tracking of cell cycle progression. Many cellular events can only occur in the sessile stalked cell including DNA replication, flagellum synthesis, and cell division.

In *CellX*, we adapt a continuum approach that accounts for phenomena involving the wholesale reorganization of structural elements within a cell. In a companion study, we implement molecular structure into an intracellular self-organization model based on a free energy functional minimization approach.

The ability of a cell to reorganize its internal structure in response to damage, changes in the host medium, division or mutation indicates the presence of several behaviors that play a key role in cellular and viral life:

- *Self-assembly*, wherein molecular-scale building blocks organize into membranes, fibrils, viruses and other structures, a process driven by the second law, that is, the tendency to minimize free energy in an iso-volumetric, iso-thermal system (Ortoleva, 2003)
- *Self-organization and other nonlinear dynamical phenomena*, wherein a system subject to far-from-equilibrium conditions (e.g., the influx of nutrients and elimination of waste products) can spontaneously develop patterns of non-uniform concentration or other variables (Ortoleva 1992, 1994)
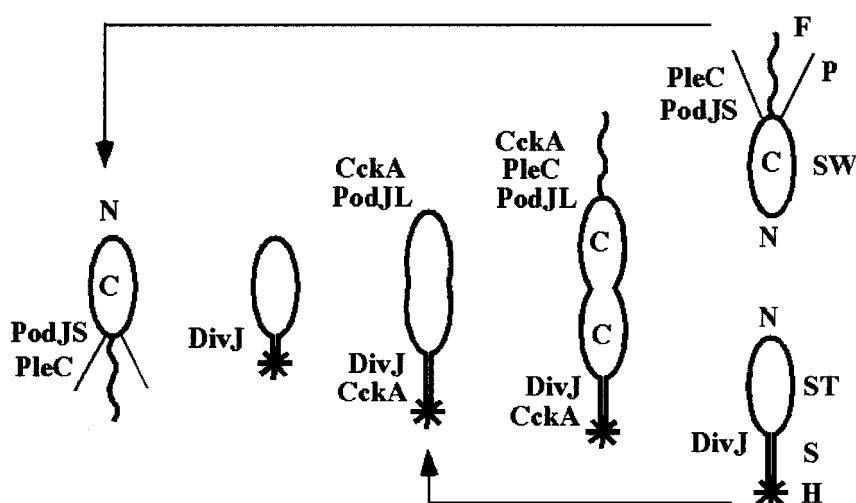

FIG. 9. Cell cycle of *Caulobacter*. Also shown is the localization of some of the major proteins that are the object of this proposal. Polar localized proteins are shown above or below the poles where they localize. PleC and CckA are distributed throughout the membrane of cells where they are not shown. DivL follows the same pattern as CckA (Newton, pers. com.). The letter C inside of cells indicates the presence of CtrA. N indicates the new pole of the cell, formed by the previous cell division. SW, swarmer cell; ST, stalked cell; F, flagellum; P, pili; S, stalk; H, holdfast.

- *Templated polymerization (notably replication, transcription and translation) and gene repair*, wherein information stored at the molecular scale can be copied or interpreted/functionalized (Weitzke and Ortoleva, 2003; Ortoleva, 2003).

To arrive at a cell model capable of accounting for these processes, one must integrate descriptive variables that capture both atomic and macroscopic quantities and the processes that underlie their dynamics. We are implementing several cell simulators that accomplish this. In essence we are developing an atomistic model of a cell that takes advantage of these situations that can be treated in an average manner (e.g., using variables such as concentrations, while other aspects of the dynamics are treated using discrete variables (e.g., all-atom descriptions). In this section our *CellX* simulator is described that allows for gradients within each intracellular compartment (an approach started by Larter and Ortoleva 1981, 1982 in the context of differentiation) (see also Ortoleva 1981a,b; Feinn 1981). The all-atom aspects of our modeling are discussed in later sections under the *VirusX®* simulator.

In *CellX*, a cell is divided into compartments labeled $\alpha = 1, 2, \ldots$ separated by membranes. Within compartment α, a set of B^α majority species have concentrations $\underline{C}^\alpha = \{C_1^\alpha, C_2^\alpha, \ldots C_{B^\alpha}^\alpha\}$ that vary continuously in space ($\vec{r}$) and time (t). Similarly, a set of B^{α^*} concentrations $\underline{c}^\alpha$ describes the spatial distribution of the B^{α^*} enzymes and other minority species in the medium within compartment α. Along the membrane separating compartments α and α', on the α side, there is a spatial distribution $\underline{\Gamma}^{\alpha\alpha'}(\vec{r},t)$ of $B^{\alpha\alpha'}$ membrane-localized molecules; $\underline{\Gamma}^{\alpha\alpha'}$ does not necessarily equal $\underline{\Gamma}^{\alpha'\alpha}$. A cell is criss-crossed by a network of fibrils to which molecules can adhere and along which they may move via diffusion or directed, active processes.

A reaction-transport cell model must reflect the duality of bulk species within a compartment and those localized to membranes or fibrils, as well as the exchange of molecules between these zones. In particular, to reflect the difference in spatial dimensionality between the bulk and the surfaces or fibrils, $\underline{C}^\alpha$ is measured in molecules/volume while $\underline{\Gamma}^{\alpha\alpha'}$ is in molecules/membrane area (and similarly, $\gamma^{\alpha n}$, the concentrations of species in compartment α along fibril n, is in molecules/distance). As a result of these dimensionalities, molecules may change character from minority to majority status with the consequent change in the timescale of their dynamics.

Key aspects of cell dynamics involves guided molecular and organelle motion along membranes and fibrils. These lower spatial dimensional domains improve the efficiency of signaling and the targeting of key components of reactions. The interplay of bulk medium and surface-localized species and their role in self-organization was explored earlier in the context of self-organization and development (Larter and Ortoleva, 1981, 1982; Ortoleva, 1981a,b; 1992). Thus a model for cellular dynamics required that accounts for the chemical physics of the exchange between, and reaction-transport within, domains of various spatial dimensionalities.

The evolution of cell models toward greater comprehensiveness should, in part, be measured in terms of the degree to which they incorporate variables describing molecular-scale features. In the context of the above description, one may set forth models that capture the dynamics of populations of molecules within (and not only those at the surface of) membranes. If these molecules are asymmetric (e.g., an ion pump), then models must account for molecular orientation (e.g., pointing from α to α' versus from α' to α). In a kinetic model, rates of transition between the above types of species (membrane- or fibril-localized or bulk) can be set forth. In what follows, we present a cellular model cast in terms of macroscopic variables (e.g., concentrations), while in the context of our virus simulator *VirusX*, we introduce the notion of mixed models which preserve key aspects of atomic-scale features.

Conservation of mass on the α side of $\alpha\alpha'$ the membrane implies

$$\frac{\partial \Gamma_i^{\alpha\alpha'}}{\partial t} = -\vec{\nabla}_\| \cdot \vec{J}_{\|i}^{\alpha\alpha'} + \sum_{k=1}^{N^{\alpha\alpha'f}} \omega_{ik}^{\alpha\alpha'f} W_k^{\alpha\alpha'f}/\varepsilon + \sum_{k=1}^{N^{\alpha\alpha's}} \omega_{ik}^{\alpha\alpha's} W_k^{\alpha\alpha's}, \; i = 1, 2, \ldots B^{\alpha\alpha'} \tag{4}$$

where

$\vec{J}_{\|i}^{\alpha\alpha'}$ = flux of membrane-bound species i along the $\alpha\alpha'$ membrane

$W_k^{\alpha\alpha'f,s}$ = rate of fast or slow $\alpha\alpha'$ reaction k

$\omega_{ik}^{\alpha\alpha'f,s}$ = stoichiometric coefficient for i in fast or slow $\alpha\alpha'$ reaction k

$\vec{\nabla}_\|$ = gradient operator parallel to the membrane

The parameter ε ($<<1$) is included to emphasize the separation in timescale between the fast and slow processes that always exists in biochemical networks.

For the B^α major components residing in the bulk, conservation of mass implies

$$\frac{\partial C_i^\alpha}{\partial t} = -\vec{\nabla} \cdot \vec{J}_i + \sum_{k=1}^{N^{\alpha f}} v_{ik}^{\alpha f} W_k^{\alpha f}/\varepsilon + \sum_{k=1}^{N^{\alpha s}} v_{ik}^{\alpha s} W_k^{\alpha s} + \sum_{k=1}^{\hat{N}^{\alpha f}} v_{ik}^{\alpha *} \hat{R}_k^{\alpha f}, \quad i = 1, 2, \ldots B^\alpha \tag{5}$$

The *R*-term is explained below. The boundary conditions on $\underline{C}^\alpha$ capture the exchange with the membrane-bound species and the transmembrane processes:

$$\vec{n}^\alpha \cdot \vec{J}_i^\alpha = \sum_{k=1}^{N^{\alpha\alpha' f}} \lambda_{ik}^{\alpha\alpha' f} W_k^{\alpha\alpha' f}/\varepsilon + \sum_{k=1}^{N^{\alpha\alpha' s}} \lambda_{ik}^{\alpha\alpha' s} W_k^{\alpha\alpha' s} + \sum_{k=1}^{N^{\alpha\alpha'}} \lambda_{ik}^{\alpha\alpha' *} \hat{W}_k^{\alpha\alpha' f *} + \sum_{k=1}^{N^{\alpha\alpha'}} \lambda_{ik}^{\alpha\alpha' t} W_k^{\alpha\alpha' t} \tag{6}$$

where $\vec{n}^\alpha$ is the local normal to the membrane bounding compartment α pointing into the interior of compartment α. The $W_k^{\alpha\alpha' t}$ are rates of transfer across the $\alpha\alpha'$ membrane, assumed positive when molecules pass to the α side. There are many transmembrane processes associated with various channels, active processes or reactions that involve species on both sides of the membrane.

The minority species must be treated carefully. By definition, their concentrations are low; we emphasize this by writing $\underline{c}^\alpha = \varepsilon\hat{\underline{c}}^\alpha$ where, as $\varepsilon << 1$, the $\hat{\underline{c}}^\alpha$ are comparable to the concentrations of the majority species $\underline{c}^\alpha$. We write their $B^{\alpha *}$ concentrations $\hat{\underline{c}}_i^\alpha$. With this, the flux of a minority species is written $\varepsilon\vec{J}_i^\alpha$. We consider the minority species to only be generated by slow reactions of rate $R_k^{\alpha s}$; however, they may facilitate cycles that replenish them (e.g., the enzyme mechanism $E + S \Leftrightarrow ES \rightarrow E + P$), all of which are fast but which establish a steady-state balance. These reactions are taken to have rates $R_k^{\alpha f}/\varepsilon$; however, they involve a factor of a minority concentration in their forward and reverse rate, so they imply a term $\hat{R}_k^{\alpha f}$ in the majority dynamics (6) where stoichiometric coefficients $\hat{v}_{ik}^\alpha$ are introduced. With this, minority mass balance implies

$$\frac{\partial \hat{c}_i^\alpha}{\partial t} = -\vec{\nabla} \cdot \vec{\hat{J}}_i^\alpha + \sum_{k=1}^{\hat{N}^{\alpha f}} \hat{v}_{ik}^\alpha \hat{R}_k^{\alpha f}/\varepsilon + \sum_{k=1}^{\hat{N}^{\alpha s}} \hat{v}_{ik}^{\alpha s} \hat{R}_k^{\alpha s} \tag{7}$$

The last term arises from rates of very slow reactions (of rate $\varepsilon\hat{R}_k^{\alpha s}$) that produce or consume the minority species (e.g., from protein or enzyme synthesis). The boundary conditions for the minority species imply

$$\varepsilon\vec{n}^\alpha \cdot \vec{\hat{J}}_i^\alpha = \sum_{k=1}^{N^{\alpha\alpha' s}} \hat{\lambda}_{ik}^{\alpha\alpha' s} W_k^{\alpha\alpha' s} + \sum_{k=1}^{N^{\alpha\alpha' f}} \hat{\lambda}_{ik}^{\alpha\alpha' f} \hat{W}_k^{\alpha\alpha' f} \tag{8}$$

Similar equations for the concentrations $\gamma^{\alpha n}$ of species localized to fibrils. Solution of the above equations yields the dynamics of the cell. As $\varepsilon << 1$, this leads to computational difficulties that are addressed in the next section.

Multiple scale analysis

The present objective of the rigorous asymptotic ($\varepsilon \rightarrow 0$) analysis of the model of the previous section is to obtain a set of well-behaved equations (e.g., ones that do not have the singular behavior in ε). As $\varepsilon << 1$, straightforward simulation is impractical due to numerical instability and the small time steps required for numerical simulation.

Our approach is similar to that used in the analysis of other multiple timescale reaction-transport systems (Ortoleva and Ross, 1975; Ortoleva, 1992, 1994; Weitzke and Ortoleva, 2003). The development here is more complex due to both the number of types of equations involved and the exchange of molecules between the bulk and the membrane surface or fibrils.

To facilitate the discussion, we recast the problem in terms of its spatially discretized form. Thus, there are sets of $\underline{C}^\alpha$ and $\hat{\underline{c}}^\alpha$ at every grid node in the interior of compartment and in the bulk close to the $\alpha\alpha'$ membrane (i.e., the boundary of the α-th continuum domain). Similarly, there are sets of $\underline{\Gamma}^{\alpha\alpha'}$ at each node along the $\alpha\alpha'$ membrane and similarly for the $\gamma_i^{\alpha n}$ along fibril n in compartment α. Let these variables for all compartment interiors, fibrils, and membranes be written as a column vector $|X\rangle$. The entire problem may be written in the symbolic form $\mathcal{A}|X\rangle + |F\rangle = \frac{1}{\varepsilon} \mathcal{V}|W^f\rangle$. Here $\mathcal{A}$ is a square matrix whose size is the same as $|X\rangle$, $|F\rangle$ is a column vector which depends on $|X\rangle$, $\mathcal{V}$ is a rectangular matrix, and $|W^f\rangle$ is a column

vector formed from all the $W_k^{\alpha f}, W_k^{\alpha\alpha' f}, \hat{R}_k^{\alpha\alpha' f}, W_k^{\alpha\alpha' s}$, and $W_k^{\alpha\alpha' f*}$ at all discretization nodal points (in all compartments and on all fibrils and membrane surfaces). To see why $W_k^{\alpha\alpha' s}$ is included, see equation (8). Note that $\mathcal{A}$ has the form $\mathcal{A}'\frac{\partial}{\partial t}$ so that it is a differential operator for the nonzero components of $\mathcal{A}'$ (e.g., boundary conditions do not involve time derivatives).

To start the analysis, introduce a set of row vectors $\langle\sigma_\ell|$ ($\ell = 1, 2, \ldots L$, L being the number of such linearly independent row vectors as determined by the structure of $\mathcal{V}$) that are orthogonal to the columns of $\mathcal{V}$:

$$\langle\sigma_\ell|\mathcal{V} = |0\rangle, \ \ell = 1, 2, \ldots L \tag{9}$$

$|0\rangle$ being the null column vector. Assuming that $|X\rangle$ has N components, this yields the well-behaved equations

$$\langle\sigma_\ell|\mathcal{A}|X\rangle + \langle\sigma_\ell|F\rangle = 0, \ \ell = 1, 2, \ldots L \tag{10}$$

Thus, $N - L$ additional equations are needed to solve the problem.

To conclude, let $\langle\xi_\ell|$ be a row vector with the same number of components as $|W^f\rangle$. By construction, $\langle\xi_\ell|$ is one of the $N - L$ independent rows of v. The $N - L$ equations needed to solve the problem take the form

$$\langle\xi_\ell|W^f\rangle = 0, \ \ell = 1, 2, \ldots N - L \tag{11}$$

These are generalized subequilibrium or steady-state conditions of the cell system. They are the generalization of similar equations arising in simpler compartmentalized cell models (Weitzke and Ortoleva, 2003). This formulation has been implemented in our *CellX* simulator (Fan et al., 2003a).

Furthermore, we are developing a reformulation of the intracellular reaction-transport dynamics that takes into account the separation of timescales between the transport of small molecules and larger entities and its interplay with fast and slow reactions. Our approach involves the introduction of a timescale ratio ε following from the ratios of rate and transport coefficients. The procedure leads to a set of modified reaction-transport equations that can be computed efficiently (Fan et al., 2003b).

Self-assembling systems and the VirusX® simulator

We are building a simulator, *VirusX*, for modeling key atomic-scale features of enzymes, viruses, and other subcellular features needed for drug discovery and to construct membranes, fibrils, and other subcellular features is a self-organizing cell model. The challenge addressed is that these features are commonly supra-million–atom objects. Thus, how can we simulate their overall dynamics (e.g., conformation changes, self-assembly), and yet preserve all the atomic-scale features as needed for drug discovery? In *VirusX*, we address this challenge via the integration of atomistic and continuum approaches.

In *VirusX*, efficient all-atom molecular mechanics simulation techniques are used that allow us to model processes that operate on the long timescale of interest. These include space-warping (Jaqaman and Ortoleva, 2002), tree codes, multiple timescale, and other methods (Jarymowycz et al., 2003). A *VirusX* simulation is seen in Figure 10. In ongoing work, we are embedding the all-atom computations in a continuum mesoscopic model to capture the host medium (bioelectrolyte or target cell surface) or virus-encasing lipid membrane or other features of lesser molecules (as in the supra-hundred million–atom HIV system).

The new version of *VirusX* presently under development is based on a set of field variables that describe the evolution in time of the spatial distribution of the position/orientation densities for a representative set of molecular species in two and three dimensions. The simulator utilizes finite element and multiple timescale techniques. The underlying equations contain mesoscopic corrections, enabling *VirusX* to mix atomic-scale and macroscopic reaction, transport and mechanical effects. When considering the carefully orchestrated activities of a cell (e.g., distribution of genetic material and organelles among progeny at cell division), one arrives at the conclusion that the cell should be conceived of as a giant, highly structured molecular association that exploits the balance of (or integrates) atomistic and macroscopic phenomena to achieve its impressive functionality and survivability. Thus, methods developed for the *VirusX* simulator will also be applied to the advanced version of *CellX*.

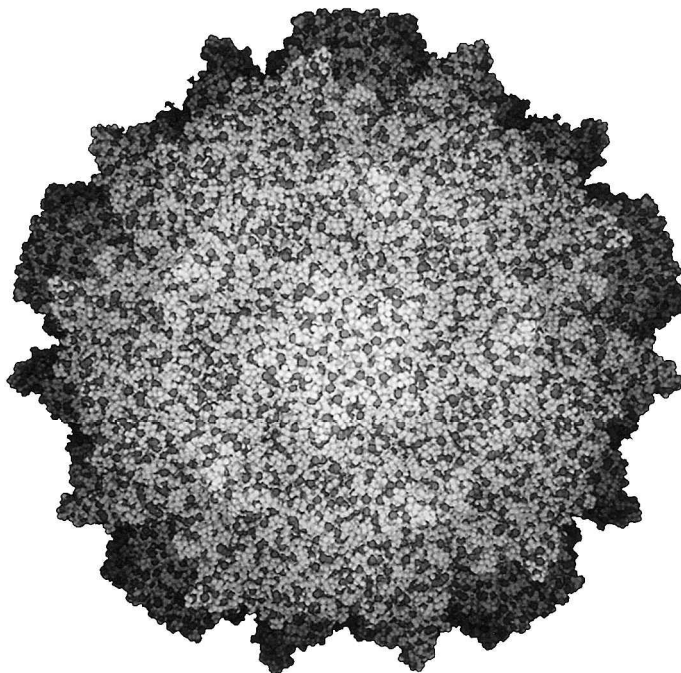For large viruses (e.g., HIV), we are developing a continuum model as follows. Much of the structure of

FIG. 10.   Three-dimensional structure of Poliovirus (Type *1* Mahoney) capsid showing 713,580 atoms in CPK shading predicted using a preliminary version of our *VirusX*® simulator (Jarymowycz et al., 2003).

the continuum description of *VirusX* follows from a free energy functional and a finite element simulation approach. Let *F* be the free energy constructed in the form

$$F = F_0 - T\Delta S + \Delta E \tag{12}$$

where F_0 is a free energy that accounts for the isotropic behavior (i.e., neglects molecular orientational order). However, F_0 contains correction terms due to the formation of large gradients associated with interfaces or the interaction of macromolecules or other bodies with the aqueous intracellular medium. ΔS and ΔE are corrections to F_0 arising from molecular orientational order and energetics. In this way, *F* is a function of the spatial distribution of the position/orientational probability density for each type of molecule in the system.

CONCLUSION

In our view of the current state of cell physiology, a comprehensive cell model must capture the coupling of the genome, proteome, metabolome and infrastructural elements as in Figure 2. We believe we have made significant advances in all these elements. We are presently completing our program through the development of *CellX* and *Karyote* for whole cell modeling, and *VirusX* for viruses and other intracellular features. In particular, we are integrating the features of *CellX* and the mixed all-atom/continuum self-assembly formulation of *VirusX* with the transcription/translation polymerization, multiple timescale methodologies and the electro-physiological formulation of the compartmented *Karyote* model. With this integration and the completion of the post-translational kinetic and gene control modules, we will have completed what we believe to be the essential structure wherein we can use gene sequence data and protein or protein aggregate structure/function rules to predict cell behavior. The calibration and augmentation of the incomplete aspects of the model with cell data is being addressed using information theory. As we have

tested all individual elements of our cell modeling approach, we believe that we are poised to deliver on the great promise that cell modeling has presented since the 1950s. Our progress can be followed at http:// biodynamics.indiana.edu.

ACKNOWLEDGEMENTS

REFERENCES

BRENNIG, L., MALEK-MANSOUR, M., and HORSTHEMKE, W. (1976). A stochastic approach to hydrodynamics. Physical Lett. A. **59A,** 341–342.

BAKKER, B.M., MICHELS, P.A.M., OPPERDOES, F.R., et al. (1997). Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. Journal of Biological Chemistry **272,** 3207–3215.

CHANCE, B., HESS, B., and BETZ, A. (1964). DPNH oscillations in a cell-free extract of *S. carlsbergensis.* Biochem. Biophys. Res. Commun. **16,** 182–187.

CHANCE, B., PYE, E.K., GHOSH, A.K., et al., eds. (1973). *Biological and Biochemical Oscillators* (Academic Press, New York).

DELLE DONNE, M., and ORTOLEVA, P. (1978). Generalized nonlinear master equations for local fluctuations: dimensionality expansions and augmented mean field theory. Journal of Stat. Phys., **18,** 319–331.

FAN, J., TUNCAY, K., and ORTOLEVA, P. (2003a). Multi-dimensional cell reaction-transport simulation: Application to division in *Caulobacter crecentis* (in preparation).

FAN, J., TUNCAY, K., and ORTOLEVA, P. (2003b). Multi-scale cellular reaction-transport dynamics: role in autonomous behavior (in preparation).

FEINN, D.M. (1981). A theoretical study of self-generated and imposed bioelectricity: the MTM model [Ph.D. thesis] (Indiana University, Bloomington).

FIRTH, C.A.J.M., and BRAY, D. (2000). Stochastic simulation of cell signaling pathways. In: Bower, J.M., and Bolouri, H., eds. *Computation Modeling of Genetic and Biochemical Networks.* (The MIT Press, Cambridge, MA), 263–286.

GILLESPIE, D.T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal Comput. Phys., **22,** 403–434.

GILLESPIE, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry **81,** 2340–2361.

GOLDBETER, A. (1996). *Biochemical Oscillations and Cellular Rhythms: The Molecular Basis of Periodic and Chaotic Behaviour* (Cambridge University Press, Cambridge).

HAHN, H.-S., ORTOLEVA, P., and ROSS, J. (1973). Chemical oscillations and multiple steady states due to variable boundary permeability. Journal of Theoretical Biology **41,** 503–521.

HAHN, H.-S., NITZAN, A., ORTOLEVA, P., et al. (1974). Threshold excitations, relaxation oscillations, and effect of noise in an enzyme reaction. Proceedings of the National Academy of Sciences USA **71,** 4067–4071.

JAQAMAN, K., and ORTOLEVA, P. (2002). New space warping method for the simulation of large-scale macromolecular conformational changes. Journal of Computational Chemistry **23,** 484–491.

JARYMOWYCZ, L., HUBBARD, V., BERRY, E., TUNCAY, K., and ORTOLEVA, P. (2003). *VirusX*®: All-atom simulation of viral structure and stability (in preparation).

JAYNES, E.T. (1957). Information theory and statistical mechanics. Physical Review **106,** 620–630.

KASCER, H., and BURNS, J.A. (1973). The control of flux. Symposium Soc. Exp. Biol. **27,** 64–104.

LARTER, R., and ORTOLEVA, P. (1981). A theoretical basis for self-electrophoresis. Journal of Theoretical Biology **88,** 599–630.

LARTER, R., and ORTOLEVA, P. (1982). A study of instability to electrical symmetry breaking in unicellular systems. Journal of Theoretical Biology **96,** 175–200.

LeNOVERE, N., and SHIMIZU, T.S. (2001). STochSim: modelling of stochastic biomolecular processes. Bioinformatics **17,** 575–576.

McADAMS, H.H., and ARKIN, A. (1997). Stochastic mechanisms in gene expression. Proceedings of the National Academy of Sciences USA **94,** 814–819.

McQUARRIE, D.A. (1976). Statistical Mechanics. (Harper & Row, New York).

MENDES, P. (1997). Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. Trends in Biochemical Sciences **22,** 361–363.

NAVID, A., and ORTOLEVA, P. (2003). Simulated nonlinear dynamics of glycolysis in the protozoan parasite *Trypanosoma brucei* (submitted to J. Theor. Biol.).

NITZAN, A., ORTOLEVA, P., DEUTCH, J., et al. (1974). Fluctuations and transitions at chemical instabilities: the analogy to phase transitions. Journal of Chemical Physics **61,** 1056–1074.

PRIGOGINE, I., and LEFEVER, R. (1968). Symmetry breaking instabilities in dissipative systems II. Journal of Chemical Physics **48,** 1695–1700.

ORTOLEVA, P., and ROSS, J. (1973). A chemical instability mechanism for asymmetric cell differentiation. Biophysical Chemistry **1,** 87–96.

ORTOLEVA, P., and ROSS, J. (1975). Theory of propagation of discontinuities in kinetic systems with multiple time scales: fronts, front multiplicity, and pulses. Journal of Chemical Physics **63,** 3398–3408.

ORTOLEVA, P., and YIP, S. (1976). Computer molecular dynamics studies of a chemical instability. Journal of Chemical Physics **65,** 2045–2051.

ORTOLEVA, P. (1981a). Electrical phenomena in developing biological systems. In: Burton, T.A., ed. *Mathematical Biology: A Conference on Theoretical Aspects of Molecular Science*, (Pergamon Press, New York), 209–224.

ORTOLEVA, P. (1981b). Developmental bioelectricity. In: Illinger, K.H., ed. *Biological Effects of Nonionizing Radiation* (American Chemical Society, Washington, DC), 163–212.

ORTOLEVA, P. (1992). *Nonlinear Chemical Waves* (Wiley, New York).

ORTOLEVA, P. (1994). *Geochemical Self-Organization* (Oxford University Press, New York).

ORTOLEVA, P. (2003). *Mesoscopic Chemical Physics* (in preparation).

PRIGOGINE, I., and LEFEVER, R. (1968). Symmetry breaking instabilities in dissipative systems II. Journal Chemical Physical **48,** 1695–1700.

RASHEVSKY, N. (1960). *Mathematical Biophysics Physico-Mathematical Foundations of Biology. Volume One*, 3rd rev. ed. (Dover, New York).

ROSS, J. (1985). Ber. Bunsenges. Physical Chemistry **89,** 605.

SAURO, H.M. (1993). SCAMP: a general purpose simulator and metabolic control analysis program. CABIOS **9,** 441–450.

SAYYED-AHMAD, A., TUNCAY, K., and ORTOLEVA, P. (2003). Towards automated cell model development through information theory. Journal of Physical Chemistry (in press).

SCHAFF, J., SLEPCHENKO, B., MORGAN, F., et al. (2001). Virtual cell. [On-line] Available: www.nrcam.uchc.edu.

SHANNON, C.E., and WEAVER, W. (1949). *The Mathematical Theory of Communication.* (The University of Illinois Press, Illinois).

SHIMIZU, T.S., and BRAY, D. (2001) Computational cell biology—the stochastic approach. In: Kitano, H., ed. *Foundations of Systems Biology.* (MIT Press, Cambridge, MA).

TURING, A.M. (1952). Chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London, **B237,** 37–72.

TOMITA, J., NAKAYAMA, Y., NAITO, Y., et al. (2001). E-cell [On-line]. Available: www.e-cell.org/.

TYSON, J. (2001). JigCell [On-line]. Available: http:gnida.cs.vt.edu/~cellcyclepse/.

VISSER, N., and OPPERDOES, F.R. (1980). Glycolysis in *Trypanosoma brucei.* European Journal of Biochemistry **103,** 623–632.

WEITZKE, E.L., and ORTOLEVA, P. (2003). Simulating cellular dynamics through a coupled transcription, translation, metabolic model (In press).

Address reprint requests to:
Dr. P. Ortoleva
Center for Cell and Virus Theory
Chemistry Building
Indiana University
Bloomington, IN 47405

E-mail: ortoleva@indiana.edu