Quantitative cell biology with the Virtual Cell☆

Boris M. Slepchenko[1], James C. Schaff[1], Ian Macara[2] and Leslie M. Loew[1]

[1]Center For Biomedical Imaging Technology and National Resource for Cell Analysis and Modeling, Department of Physiology, University of Connecticut Health Center, Farmington, CT 06030, USA
[2]Center for Cell Signaling and Department of Pharmacology, University of Virginia, Charlottesville, VA 22908, USA

Cell biological processes are controlled by an interacting set of biochemical and electrophysiological events that are distributed within complex cellular structures. Computational models, comprising quantitative data on the interacting molecular participants in these events, provide a means for applying the scientific method to these complex systems. The Virtual Cell is a computational environment designed for cell biologists, to facilitate the construction of models and the generation of predictive simulations from them. This review summarizes how a Virtual Cell model is assembled and describes the physical principles underlying the calculations that are performed. Applications to problems in nucleocytoplasmic transport and intracellular calcium dynamics will illustrate the power of this paradigm for elucidating cell biology.

Fundamental cell biological processes, such as growth, signaling, differentiation and death, are extraordinarily complex, depending upon an enormous number of different molecules and molecular interactions. High-throughput methods for analysis of DNA and RNA (genomics) and protein (proteomics) are now providing an overwhelming amount of qualitative data, cataloguing the molecules involved in these complex pathways. The field of 'systems biology' [1–3] uses computational analyses to sift through the large volume of data, to identify unique sets of molecules involved in particular cellular responses, and to develop statistical relationships between molecules that suggest causal interactions. Such largely qualitative data and analyses can offer significant insights into the operation of large, complex biochemical systems, such as those involved in metabolic and signaling networks.

To develop a full understanding of the mechanisms underlying a cell biological event, however, the quantitative physical–chemical details must be elucidated. As a first prerequisite, this requires measurements of the concentrations, locations, biochemical reaction rates and membrane transport kinetics for the involved molecules. Such data can come from traditional biochemical methods, but are also available through direct *in vivo* measurements on live cells, using modern microscope imaging with fluorescent probes and indicators and electrophysiological techniques such as patch-clamp single-channel recording. Once such data are acquired, a crucial and obvious challenge is to determine how these, often disparate and complex, details can explain the cellular process under investigation. The ideal way to meet this challenge is to integrate and organize the data into a predictive model. Indeed, the value of computational approaches in cell biology has been increasingly recognized as demonstrated by several recent reviews [4–6] and a book [37]. However, for cell biologists, formulating a mathematical model based on quantitative data and solving its equations to generate simulations can present significant technical challenges. Overcoming this barrier, in order to complete the cycle of hypothesis–prediction–experiment, is required to understand the biophysical mechanisms underlying the function of a cellular subsystem. To meet these needs for quantitative cell biology, we have developed a software-modeling environment called the Virtual Cell. [7–9].

The Virtual Cell provides a formal framework for modeling biochemical, electrophysiological, and transport phenomena while considering the subcellular localization of the molecules that take part in them. This localization can take the form of a three-dimensional (3D) arbitrarily shaped cell, where the molecular species might be heterogeneously distributed. The geometry of the cell, including the locations and shapes of subcellular organelles, can be imported directly from microscope images. Such a model explicitly considers the diffusion of the molecules within the geometry, as well as their reactions. Simulations from such a 'spatial' model require the solution of partial differential equations. Alternatively, the localization can be more crudely represented by considering the average concentrations of molecules within well-mixed compartments that correspond to the cellular and subcellular structures. This corresponds to the assumption that diffusion is fast compared with the timescales of all the biochemical transformations, so that any diffusible molecule will be uniformly distributed as soon as it is produced and no matter how convoluted the geometry of its compartment; such a nonspatial model will result in ordinary differential equations that can be numerically solved much more efficiently. Thus, these two classes of models result in very different mathematical forms, although the original quantitative mechanistic hypothesis could be the same. Simulation results on the dynamic distributions of the molecules controlling a cellular function are displayed

☆ Supplementary data associated with this article can be found at doi: 10.1016/j.tcb.2003.09.002

Corresponding author: Leslie M. Loew (les@volt.uchc.edu).

 0962-8924/$ - see front matter © 2003 Elsevier Ltd. All rights reserved. doi:10.1016/j.tcb.2003.09.002

Box 1. Access to the Virtual Cell and its database

The Virtual Cell software is maintained at a central server within the National Resource for Cell Analysis and Modeling (NRCAM) at the University of Connecticut Health Center, and is available to users worldwide through JAVA technology on a web browser. This web-based client-server architecture has several advantages: it obviates the task of software distribution and installation on heterogeneous hardware platforms – the Virtual Cell runs equally well on Windows, Linux and Macintosh machines; enhancements and bug fixes are available to the entire user community as soon as they are deployed; the computationally intensive calculations, which might be unacceptably slow on the computer of a user, are optimized for a dedicated cluster of fast computers at NRCAM; and, most importantly, a central database of user models, maintained at NRCAM, facilitates the online sharing of model components, collaboration, and interactive publication of models.

within the software environment and can also be exported as arrays of numerical values, sets of images or digital movies. Box 1 provides details of the software and how to access it through the Internet.

Quantitative cell biology

The paradigm that we term quantitative cell biology is described in Figure 1, where a study of calcium dynamics in a neuronal cell [10,11] is used as a concrete illustration. In this study, the neuromodulator bradykinin applied to the cells produced a calcium wave that starts in the neurite and spreads to the soma and growth cones. The calcium wave was monitored using digital microscope imaging of a fluorescent calcium indicator. The hypothesis was that interaction of bradykinin with its receptor on the plasma membrane induced the production of inositol (1,4,5)-trisphosphate (Ins(1,4,5)P_3) that diffused to its receptor on the endoplasmic reticulum (ER), leading to calcium release. The model contained details of the relevant receptor distributions (through immunofluorescence analysis) within the cell geometry, the kinetics of Ins(1,4,5)P_3 production (through biochemical analysis of Ins(1,4,5)P_3 in cell populations and photorelease of caged Ins(1,4,5)P_3 in individual cells); the transport of calcium through the Ins(1,4,5)P_3 receptor calcium channel and the SERCA pump (from literature studies of single-channel kinetics and radioligand flux); and calcium buffering by both endogenous proteins and the fluorescent indicator (from confocal measurements of indicator concentrations). The mathematical equations generated by this combination of molecular distributions and reaction and membrane transport kinetics was then solved to produce a simulation of the spatiotemporal pattern of calcium that could be directly compared with the experiment.

The procedure described in Figure 1 is just a restatement of the scientific method, where 'model' is simply defined as a complex quantitative 'hypothesis'. Practically, however, a model is typically composed of several hypotheses, where even the choice of which of the hypothetical components to include can in itself be a hypothesis. Because of the complexity of models, it is difficult to formulate them as a string of verbal hypotheses; a text format also cannot capture the quantitative relationships. Model features are, therefore, most effectively expressed in terms of diagrams linked to mathematical expressions. Figure 1a shows diagrams taken directly from the Virtual Cell user interface for the compartmental distribution of the molecules, transport processes in the ER membrane, and the assignment of compartments to the spatial geometry of a neuroblastoma cell. Furthermore, generating quantitative predictions from a model requires solving mathematical equations (i.e. simulations). Computational tools such as the Virtual Cell facilitate both the process of model construction and the generation of simulations.

Building a model in the Virtual Cell

Building a biological model in the Virtual Cell is an iterative process that should start with a simple model that is incrementally validated against simulations and experimental observations. In our calcium dynamics illustration, for example, an initial assumption was that the Ins(1,4,5)P_3 receptor was distributed uniformly throughout the cell. The simulations produced under this assumption gave calcium amplitudes that were much too high in the neurite and too low in the soma compared with the experimental calcium dynamics. This prompted us to investigate the Ins(1,4,5)P_3 receptor distribution by immunofluorescence, which revealed that the calcium density in the soma was twice as high as that in the neurite; the simulation resulting from this new elaboration of the model provided an excellent match to the experimental calcium wave. Furthermore, a model (or hypothesis) should lead to predictions that can be further tested with new experiments. For example, our model predicted that if bradykinin were applied to different regions of the cell, rather than globally, application to the proximal neurite would be sufficient to produce a calcium wave; but activation of bradykinin receptors in only the growth cone or soma would produce an aborted calcium wave. These predictions were subsequently borne out by experiments in which bradykinin was regionally applied through pressure application, using a micropipette [10].

A Virtual Cell biological model is composed of three parts in an expanding tree structure: a single 'physiological model' that captures the mechanistic hypothesis; one or more 'applications', where experimental conditions, geometry and modeling approximations are introduced to pose a concrete mathematical problem; and one or more 'simulations', where each represents a numerical solution to the mathematical problem posed by an application (Figure 2).

A physiological model consists of molecular species and mechanisms (reactions and fluxes) localized to cellular structures. Cellular structures are created by specifying the topological arrangement of membranes and membrane-bound compartments. Biochemical reactions are defined within volumetric compartments of the cell, as well as in membranes; molecular fluxes and electrical currents are defined across membranes. The reaction and flux rates are determined as an explicit function of the local environment (e.g. concentrations, surface densities and membrane potential) and one or more kinetic parameters (e.g. k_f and k_r, the forward and reverse mass action rate

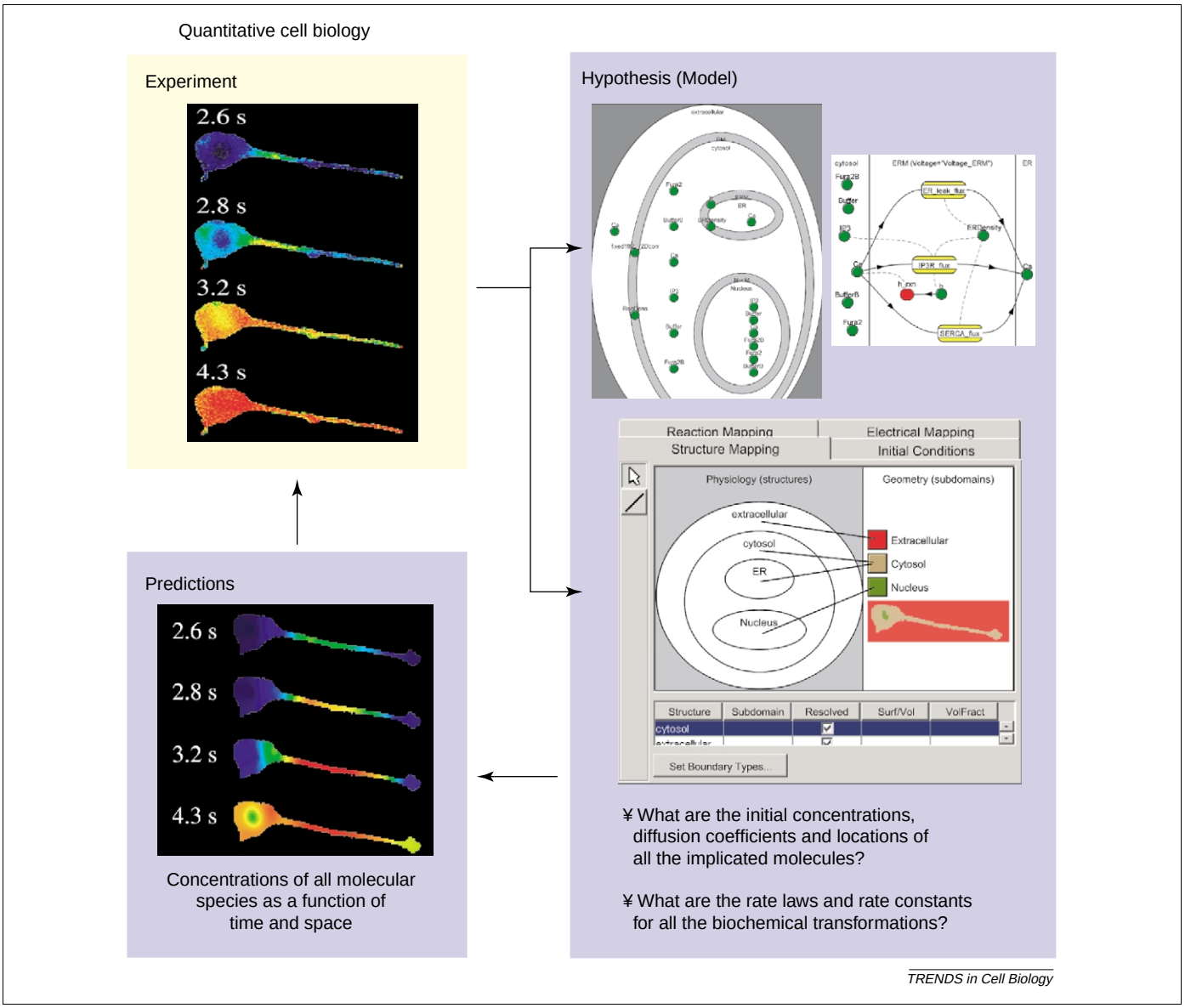

Figure 1. Quantitative cell biology is dependent on computational modeling. Experiments provide image, biochemical and electrophysiological data on the initial concentrations, kinetic rates, and transport properties of the molecules and cellular structures, which are presumed to be the key components of a cellular event. If the system is perturbed by addition of a ligand, electrical stimulus or other experimental intervention, the model should be capable of predicting changes in the spatiotemporal distribution of all the molecules. This cycle is illustrated by reference to a study of calcium waves evoked by activation of the bradykinin receptor in the plasma membrane of a neuronal cell. [10,11] Inputs to the model in the Hypothesis panel include the identity of the molecules hypothetically involved in calcium release, and the cellular compartment in which they are located (upper left); the reactions and transport processes between theses molecules (illustrated with a scheme for transport and reactions within the endoplasmic reticulum (ER) membrane); the assignment of compartments and molecular locations to a specific cell geometry (illustrated with the geometry of a neuroblastoma cell in the center); and specification of initial concentrations, boundary conditions and diffusion coefficients. This information is sufficient to generate simulations of the calcium dynamics (illustrated in the Predictions panel) that can be directly compared with the experimentally observed calcium dynamics (illustrated in the Experiment panel). If the prediction and the experiment do not agree, the assumptions underlying model inputs will need to be reexamined. The model can also be used to perform 'virtual' experiments in advance of a real experiment and can simulate the behavior of molecules that are not easily visualized experimentally.

constants). Mass action and Michaelis–Menten rate laws are available automatically, but user-defined general kinetic expressions are also possible. Membrane transport kinetics can be specified with expressions for molecular flux or, for ions, the electric current. The transport kinetics can be described in terms of standard electrophysiological formulas (e.g. Goldman–Hodgkin–Katz permeability or Nernst conductance) or as user-defined molecular flux or current.

After the physiological model has been defined, a non-spatial application should be constructed to investigate the steady-state (basal) behavior of the model. A steady state is achieved when concentrations and membrane potentials do not change over time. If the relative surface areas and volumes of the cellular structures are specified, based on the true 3D cellular geometry, the steady state compartmental solution will generally also apply to the full spatial problem. It is worth noting that most physiological models are not 'closed' systems, and that they will exchange mass and energy with the environment. For example, in a typical physiological state a model of glucose metabolism would continuously consume glucose, oxygen and ADP and

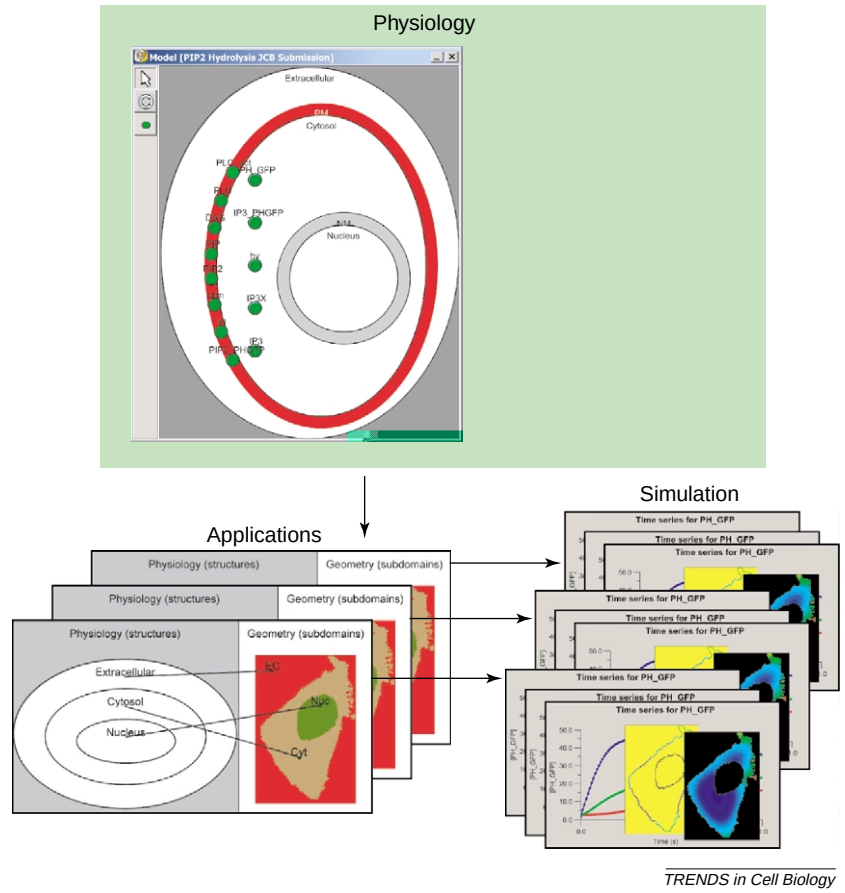

Figure 2. A Virtual-Cell biological model consists of a single mechanistic physiological model that encodes the general structure and function of the cellular system, and multiple applications that allow specific geometries, initial concentrations, protocols and modeling assumptions required to apply the model to various experimental contexts. Each application, in turn, can spawn multiple simulations with different numerical methods and varying parameters.

continuously create ATP and carbonic acid. If this steady state consumption and production were not considered, the model would consume all metabolites and not achieve a physiological steady state. Therefore, the model boundaries are set with either a fixed concentration (equivalent to an infinitely large bathing medium) or a fixed rate of flux across the boundary, which, when set to zero, corresponds to a closed barrier.

Next, the response of the system to external stimuli might be considered. Typical perturbations include: hormone injected into the extracellular media, voltage- or current-clamp protocols, and photorelease of a caged compound within the cytosol. It is essential, however, that the steady-state behavior be characterized first, so that the response to the stimuli can be distinguished from a spontaneous change caused by nonequilibrium initial conditions.

After characterizing the steady state and transient behaviors of the nonspatial system, one or more spatial applications can be created to investigate the impact of cell geometry. Spatial simulations are necessary to capture the behavior of the model when the distributions of molecular species are spatially inhomogeneous and, on the timescale of the biochemical transformations, diffusion is slow. A cellular geometry must be defined based on either idealized shapes (e.g. spheres, hemispheres and cylinders), or on a segmented experimental image that has been classified according to the cell morphology (e.g. an image or image stack with one color for cytosol, another for extracellular, and yet another for the nucleus as in Figure 2). Inhomogeneous initial conditions might be specified as an explicit function of spatial coordinates. Boundary conditions are then specified to determine how the diffusive transport will behave at the edge of the simulation domain. This realistic spatial representation allows for the direct comparison of simulation results with data obtained from tools such as fluorescent microscopy. This technology allows modeling to investigate the impact of cell morphology and localization on cellular physiology.

An application is sufficient to fully specify the mathematics to be passed to a simulation. As fully described in the next section, several simulations, using different numerical solver algorithms and algorithmic parameters (e.g. time steps and spatial mesh resolutions), can be generated from a single application. Model parameter variations and sensitivity analysis can also be performed at the level of the simulations. Results are displayed in a convenient window that allows for visualization of temporal and spatial variations of any concentration, flux,