

Intracellular Signaling: Spatial and Temporal Control

Ion I. Moraru and Leslie M. Loew

Center for Cell Analysis and Modeling,
Department of Cell Biology,
University of Connecticut School of Medicine,
Farmington, Connecticut
moraru@panda.uhc.edu

Cells integrate many inputs through complex networks of interacting signaling pathways. Systems approaches as well as computer-aided reductionist approaches attempt to “untangle the wires” and gain an intimate understanding of cells. But “understanding” any system is just the way that the human mind gains the ability to predict behavior. Computer simulations are an alternative way to achieve this goal—quite possibly the only way for complex systems. We have new tools to probe large sets of unknown interactions, and we have amassed enough detailed information to quantitatively describe many functional modules. Cell physiology has passed the threshold: the time to begin modeling is now.

As cell biology entered the 21st century, a convergence of factors promised to revolutionize the way state-of-the-art research is done. On the one hand, there has been an explosion of new data: some from “-omic” technologies and other high-throughput methods (but often of unknown relevance and quality, and sometimes even difficult to catalog in a consistent way) and some from new imaging and *in vivo* technologies that can reveal molecular interactions in a real and quantitative way (but often difficult to analyze or to extend to large scales). One common theme is the fact that to properly leverage these data, we need computers to process them and to help us use them for generating and testing hypotheses. Traditional bioinformatics is being transformed and complemented by the expanding systems biology and the emergent computational cell biology (6, 29, 32, 55). Many different approaches exist, ranging from the very abstracted, high-level methods of inferring connectivity maps to the very specific, mechanistic, kinematic methods of detailed quantitative spatial simulations. Which of these approaches are most useful is an open debate—and the answer likely depends on the specific system studied and the specific goal of the biologist/modeler. It is clear, however, that we will need both to refine all of these methodologies and to develop effective strategies to connect them and share data between them to fully unravel the complexity of intracellular molecular networks (30). The latter is a prerequisite for the complete determination of biological systems—a lofty goal, but one that, in view of the current progress in biological investigative techniques, mathematics, and computer science, may be achievable over the long term (54).

This review explores the promises and pitfalls of computational modeling in cell biology, focusing on paradigms and examples from the world of

intracellular signaling, with special emphasis on the role of morphology and spatial simulations. However, the scope of intracellular signaling networks is increasingly difficult to delineate, given the multiple molecular interactions present inside the cell. Thus it is not clear where to draw the boundary between signaling networks on the one hand and either metabolic networks or gene-regulatory networks on the other. We do not attempt to make such clear distinctions here, and we will mainly use the more global view of modeling and simulating intracellular molecular networks. We conclude that the field of cell biology is ripe for widespread use of computational approaches and that the best avenue for progress is to combine top-down, systems-level modeling with bottom-up, detailed quantitative modeling.

We will review both of these classes of methodologies (with more detail and emphasis on the second) in an attempt to guide the reader as to their relative roles and appropriateness for tackling particular modeling needs related to intracellular signaling.

The Top-Down View: Reverse Engineering Molecular Networks

More than three decades have passed since the measurements of changes in cytosolic cAMP and free Ca^{2+} concentrations in response to agonists introduced the concept of second messengers and intracellular signaling pathways. Until recently, these pathways have been mostly studied in the typical reductionist fashion, with researchers trying to isolate and characterize each step, from receptor activation to final effect. A pervasive assumption was that these are literally pathways, *i.e.*, that they propagate information linearly (amplification cascades notwithstanding), and that

alternate pathways generally interact at the beginning (e.g., a receptor that activates or inhibits more than one pathway in the same cell) or at the end (e.g., synergistic or antagonistic effects of kinases belonging to different pathways on the final effector proteins) but rarely at the level of the many intermediate steps. However, data from traditional biochemistry and molecular biology as well as the systematic data gathering of genomics and proteomics has shown an ever-increasing number of participants in these pathways and a much more complex and subtle interaction between them, and thus we will need computational approaches to properly resolve them (41).

Recently, as “signaling pathways” became “signaling networks,” a flurry of interest in applying network theory and related techniques to the study of intracellular signaling developed. The premise that the analysis of a complex network as a single large functional ensemble will allow us to infer its emergent properties and behavior is very attractive.

A classical example of a well-developed map of intracellular molecular networks is the metabolic network. Mathematical and engineering tools have been applied to the study of metabolism, and they have been very successful, both at prediction of the behavior of the system and at identification of biological motifs and modules (63). Universal principles such as robustness and control mechanisms have been extracted, and metabolic control analysis has become a very well established methodology for quite some time, with its own set of laws and practical applications (15). However, at present, studies of metabolism are not a good template for

“The premise that the analysis of a complex network as a single large functional ensemble will allow us to infer its emergent properties and behavior is very attractive.”

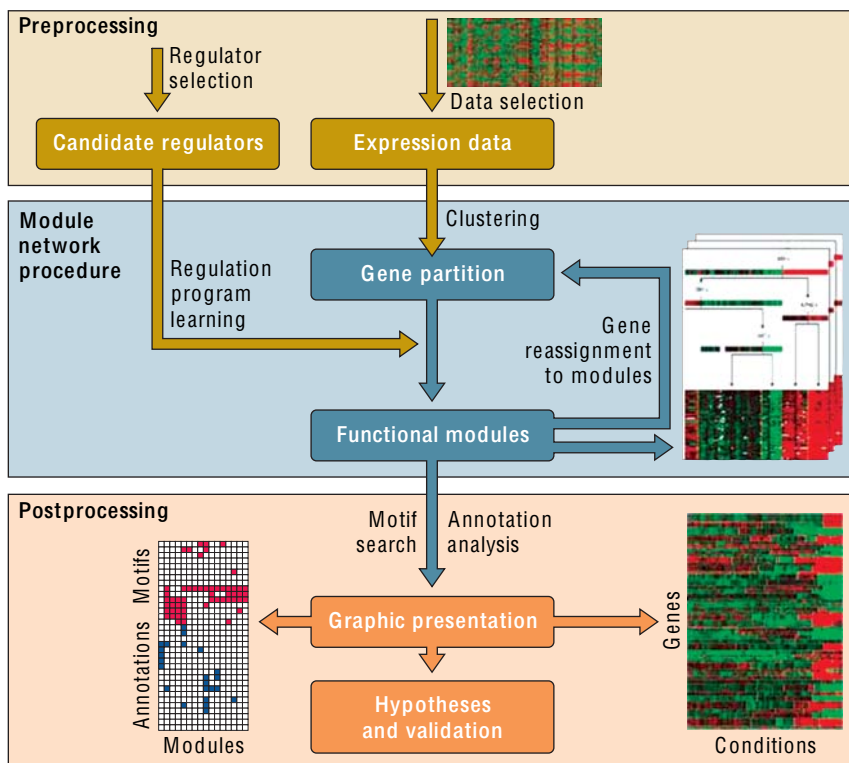
approaching signaling (and gene) networks. As a result of decades of detailed biochemical studies, the individual molecules participating in the metabolic network are very well characterized, and essentially all links between the nodes are known, in most cases, at a very good level of quantitative detail (rate laws, binding, and regulation parameters). Although a few subcomponents of signaling pathways are characterized to a similar extent—good examples are Ca^{2+} signaling in cardiomyocytes (23, 52) and EGF receptor signaling in fibroblasts (51, 62)—many of them are not, and a

large number of interactions (links in the network) are not known at all; this lack of knowledge is even more prevalent for gene-regulatory networks. Additionally, metabolic networks are probably at least an order of magnitude smaller in terms of numbers of components and interactions.

It is almost a cliché to say that the most important characteristic of a network is its architecture, but what does architecture actually mean? It is the collection of features present in the map of the network. Therefore, one of the current challenges in the case of signaling networks is to first discover the actual map of the network (1, 10). The question of what exactly to include (genes? proteins? small molecules? all of the above?) and at what granularity (modules? whole cell?) is subject to debate. Traditionally, network maps have developed slowly as the result of painstakingly combining and reconciling the results of many separate studies of individual molecules and their intracellular functions and interactions. However, given the large number of components involved and the high level of interactivity, and given the recent availability of high-throughput experimental data for both molecular interactions and functional analyses, deciphering maps of intracellular networks has recently fallen under the domain of systems biology and computational biology.

One school of thought advocates the use of statistical methods developed in the field of complex systems studies (31). These methods promise that we can infer the structure of the network, i.e., “learn the model,” with little detailed and accurate information about all of the nodes and their interaction. Large-scale datasets appropriate for this purpose are being generated by the advent of recent high-throughput technologies (34). This approach has been named alternatively as inference, system identification, or reverse engineering. The mathematical tools that have proven to be very useful for this purpose are probabilistic graphical models, such as Bayesian networks, Markov networks, or chain graphs (18). The details of applying these methods to biological network reverse engineering is beyond the scope of this review. However, we must mention that, particularly in the study of gene networks and genetic regulatory mechanisms, there have been a number of recent examples where this approach was successfully applied to identify new regulatory links or to infer modules of related genes from microarray expression data; a prototypical example (53) is illustrated in **FIGURE 1**. The individual network motifs and critical nodes identified would become the subject of the traditional experimental research, which is often needed to eventually separate epiphenomenon from causation (e.g., when is coexpression actually coregulation?). Once new hypotheses are

A Strategy for network inference and module characterization



B Sporulation and cAMP pathway module

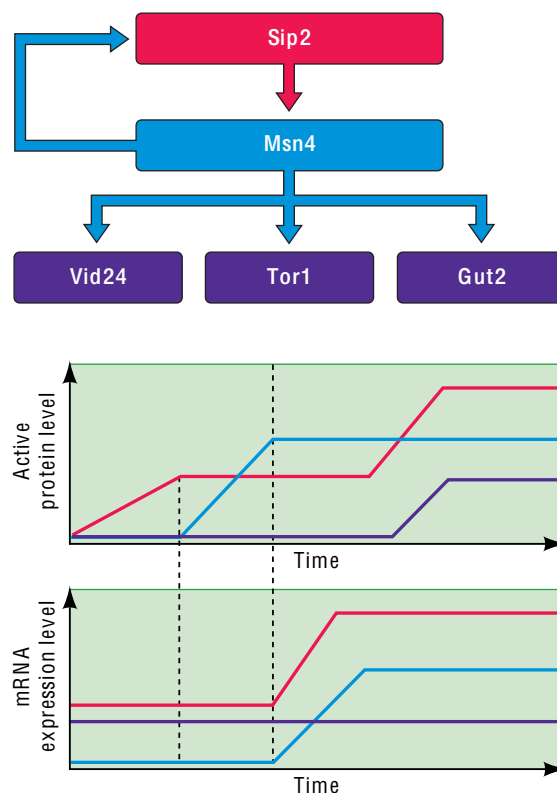


FIGURE 1. An example of network inference and module characterization from *S. cerevisiae* gene expression data sets

A: an automated module network procedure takes as input a gene expression profile data set and a precompiled set of candidate control genes. The procedure is iterative and determines both the partition of genes to modules and the regulation program for each module and is followed by testing for enrichment of gene annotations and *cis*-regulatory binding site motifs. B: sample module identified [top, a positive signaling loop involving a signaling molecule (Sip2), a transcription factor (Msn4), and targets] and the dynamic behavior encoded by it (bottom). The coordinated expression changes in signaling molecule and targets allow signaling molecule (but not transcription factor) to be correctly inferred as regulator. Figure adapted from Ref. 53, with permission (<http://www.nature.com>).

being tested, the confidence in the network structure grows and the model can be refined by targeted experiments using perturbation analysis. Additionally, once the map is known with some certainty, perturbation analysis can be even more useful, since it can generate additional testable predictions of network behavior without the intimate quantitative knowledge of individual interactions (20).

However, top-down network inference usually fails in the case of large, heavily interconnected networks. Fortunately, intracellular networks, although at first glance appearing to fall into this category, in reality are rather sparse, i.e., the average number of interactions of any given node is significantly smaller than the total number of nodes in the network (e.g., an individual kinase may phosphorylate a large number of targets, but this number is small compared with the thousands of expressed proteins in a human cell). Besides being sparse, another common feature of intracellular networks is that the number of connections per node does not have a normal (Gaussian) probabilistic distribution. Rather, a few select nodes exhibit large numbers of connections, but the

majority have a much smaller number of connections, forming a so-called scale-free network (7). The highly interconnected nodes can be viewed as “common currencies” or “hubs” through which the flow of information does fan in and out in recurring bow-tie patterns (11).

Reductionism Redux: The Devil is in the Details

During a century of cell biology studies, many new techniques have been added to the arsenal of the expert researcher; a few of them are quite revolutionary, and some of the new computational modeling approaches promise to fall into this latter category. However, the road to success has always followed the paradigm of applying the right tool for the problem at hand, not allegiance to the fad of the day or to a particular camp promoting a singular technology. So what does the cell physiologist bent on computational modeling learn from these high-level approaches? And when should we turn to different, more quantitative methods? The first subsection below analyzes some of the insights revealed by the quest to untangle connectivity

maps through network inference studies. The following four subsections detail some of the specific contexts of cell physiology that often call for shifting the focus on quantitative modeling and simulation of individual subsystems. We identify three layers of complexity in signaling networks that make accurate characterization and predictive simulations difficult when high-level modeling approaches are used exclusively (stochastic variability, dynamic evolution, and spatial distribution), and we also show examples of using complementary methods to overcome such difficulties. These examples illustrate some of the multifaceted approaches available today to the savvy modeler.

Network modules

A common feature of intracellular networks is that the connections between nodes are neither random nor regular. Often, there are numerous localized interactions among groups of nodes and few distant connections traversing many such groups of nodes, thus forming functional “modules.” Such large, sparse, scale-free, and modular networks are common in natural or designed complex systems (e.g., the market economy, the internet, VLSI chips), but they are not ubiquitous (e.g., neural nets, either biological or artificial, are neither sparse nor modular). However, metabolic, signaling, and gene-regulatory networks are most likely to turn out to be in all cases sparse, scale-free, and modular, due to the fact that such networks are best suited to balance adaptability and robustness (24, 58), and thus this is the most likely architecture in the evolutionary design of systems that are critical for species survival. Detailed studies in many different organisms of the best-characterized intracellular network, the metabolic network, have confirmed the ubiquity of a pattern of hierarchical modularity (50). It is therefore appropriate to ask

whether we should rather focus on the detailed analysis, modeling, and validation of the individual modules, which can be then used as building blocks for more comprehensive quantitative models that could eventually become very accurate.

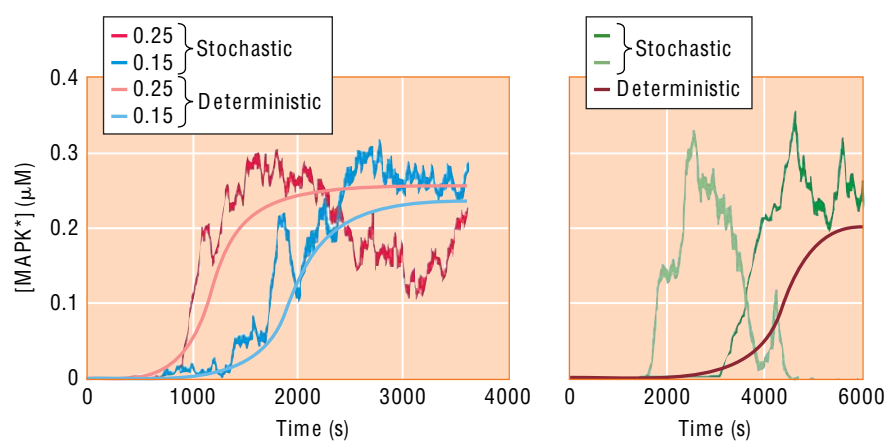
This is in essence the traditional reductionist approach to science, and the application of mathematical and computational tools to characterize individual modules has been very successful. Some functional modules can be very simple and are recognizable after a cursory look at a wiring diagram, such as the well-known feedback or feed-forward loops. However, experiments have shown that, even for such simple modules, careful mathematical analysis is necessary to predict the signal-processing characteristics of the module (the dynamic behavior encoded by it), because the quantitative details of the kinetic parameters governing the interactions usually make the system highly nonlinear. For example, a negative feedback loop can generate either homeostasis (adaptation) or an oscillatory response, and a positive feedback loop can generate hypersensitivity (thresholding) or bistability (5, 45).

It is therefore more appropriate to characterize a module not by its wiring principles but rather by the type of response it generates. A simple analogy is that in an electrical circuit it is more important to know whether a specific circuit component functions as a transistor or as a diode than to know whether it is a semiconductor or a vacuum tube (35). When building such a “parts list” of a cellular network, we find repeated occurrences of a relatively small number of different types of modules, among the most common types being amplifiers, attenuators, oscillators, “push buttons” (when stimulus needs to be maintained to maintain the response), “toggle switches” (when response remains stable after the stimulus disappears—often

FIGURE 2. An example of stochastic effects on network module dynamics

A positive-feedback loop involving MAPK signaling in Purkinje dendrite spines was modeled and simulated using either deterministic or stochastic methods. The module exhibits bistable behavior for MAPK concentration when simulated deterministically, switching from a low steady state to a high steady state after a suprathreshold stimulation by an increase in cytosolic Ca^{2+} . The small volume of the spine (1 fl) exposes the stochastic effect due to the very low number of molecules involved. The magnitude of this effect depends on diffusion coefficient and stimulus strength. If mobility of the molecules is low ($D = 0.001 \mu\text{m}^2/\text{s}$, $\text{Ca}^{2+} = 0.15 \mu\text{M}$; left), stochastic runs approximate the deterministic simulations quite well at low stimulus values ($\text{Ca}^{2+} = 0.15 \mu\text{M}$) but show increasing deviations as the stimulus is increased ($\text{Ca}^{2+} = 0.25 \mu\text{M}$), mostly as prominent fluctuations at steady state.

If mobility is higher ($D = 0.01 \mu\text{m}^2/\text{s}$; right), the behavior of stochastic simulations is completely altered, showing wildly differing patterns among separate runs with the same parameter values (including possible oscillations between the low and high MAPK concentrations in lieu of a final steady state). Figure adapted from Ref. 4, with permission.



responsible for irreversible responses and checkpoint controls), and “sniffers” (short-lived response only when stimulus changes in intensity—often responsible for adaptation/desensitization) (60). Predictive models (based on the detailed understanding of the underlying biology and mathematics) of such basic functional modules allow us to choose appropriate strategies to target control mechanisms in the network, and these models raise the prospect of being able to alter the network architecture by rewiring or constructing new modules (25). Recent proofs of concept have been the successful design and expression of synthetic genetic toggle switches and oscillators (12, 19).

Stochastic variability

One problem that we often face with generic models of intracellular networks is created by the fact that many molecules in the network are present in low copy number—this is especially true of mRNA and proteins but also of small molecules (e.g., in a volume of 0.5 μm diameter, which is comparable with the size of the lumen of a mitochondrial crista invagination or to the area surrounding a sarcoplasmic reticulum cistern, a concentration of 100 nM Ca^{2+} turns out to be only 4 Ca^{2+} ions). Some of the methods discussed above employ probabilistic techniques and thus can be adapted to capture noise and stochasticity, but the problem lies in the fact that individual modules can dramatically alter their encoded behavior when only small numbers of molecules are involved. **FIGURE 2** shows examples (3, 4) where such effects can be easily grasped intuitively. If a module that exhibits bistability operates with small numbers of molecules so that stochastic noise is significant, random fluctuations can overcome hysteresis; instead of bistability, such a system will oscillate between the normally stable steady states according to some function of the noise frequency! Similarly, large stochastic fluctuations will alter the behavior of a module that encodes hypersensitivity by increasing the sensitivity and decreasing the sharpness of thresholding up to the point that, when measured in large populations, hypersensitivity is transformed into a graded response. Therefore, in such cases we must resort to kinetic models using Monte Carlo simulations or to stochastic differential equation approximations to properly describe the system (see the references for the examples above and also the section on methods and tools below).

Dynamic behavior

A second problem is time course and evolution. Temporal regulation and control is often hard to extract using Bayesian networks or similar methods. A simplistic argument is that molecular interaction data can be meaningless for a pair of pro-

teins that are not expressed at the same time during the cellular process studied. To account for this, recent studies have used time series data to help identify a causal relationship during reverse engineering of network structures (44). Additionally, other mathematical approaches have been developed to predict dynamic systems behavior without necessarily knowing from the beginning the exact makeup of the system. These are fundamentally different from the global network inference studies because they focus on the importance of actual quantitative descriptions of the system embodied by the numerical parameter values inside the model. Ideally, kinetic descriptions of individual components should be translated into systems of differential equations (deterministic or stochastic) that can be solved numerically to predict the systems' evolution over time. However, inaccuracies can easily compound when simulating temporal changes and can decrease the predictive power of the model. Therefore, modelers have searched for general principles to sort out the complex dynamic behavior of realistic models of regulatory networks, using dynamical systems theory. Kinetic equations may seem like a good starting point, since they actually predict the evolution of the system, but often they have a huge parameter space that defies optimization. Moreover, for the understanding of the system, the perfect fitting of kinetic equation parameters can sometimes be less important than characterizing the properties of a particular state, e.g., whether it is resistant or not to perturbations or what specific state variables (e.g., component concentrations) or parameter values are critical to its projected future. When applying kinetic equations to a state of the system, we define a vector field in state space that determines the temporal behavior of the system. Vector fields can be analyzed mathematically to identify important properties such as stable and unstable steady states, checkpoints, attractors, and limit cycles, using tools such as bifurcation diagrams. A good example are studies of the cell cycle, in which the analysis of initial, mostly theoretical models (47) was able to capture many aspects of *in vivo* and *in vitro* behavior and to suggest critical experiments, which eventually led to the development of almost complete quantitative models able to predict not only normal behavior but also that of mutant phenotypes (9).

Spatial models

The third problem arises from the fact that in many cases we don't deal with a simple homogenous bag of molecules. Even prokaryotes, which usually are devoid of compartmentalization, appear to have numerous supramolecular complexes that define regions of specialized function. Calculations from first principles of molecular collision probabilities

due to normal random motion in volumes on the order of $1 \mu\text{m}^3$ or larger suggest that restricting the spatial mobility of enzymes may be the only mechanism to achieve reaction rates that are not close to zero (A. Slepoy, Sandia National Laboratory, personal communication). Additionally, intracellular concentration gradients and spatial organization of molecular complexes can play a significant role in regulating microbial activities such as chemotaxis and quorum sensing and are more common than originally thought (40, 57).

Mammalian cells have multiple internal compartments, and the possibility of interactions between different molecules is tightly controlled. Processes such as translocation of proteins to and from membranes are commonplace, and fluxes across membranes need to be specifically taken into account in many models. Internalization and endocytosis of surface receptors frequently play an important role in shaping the response to signals; EGF signaling is a good example (26). An extreme example of exquisite spatial organization that regulates signaling is found in cardiomyocytes. There, excitation-contraction coupling and force generation is characterized in good quantitative detail and has seen many successful comprehensive modeling efforts across different scales (these are beyond the scope of this review and have been recently discussed in the first issue of the new *Physiology*; see Ref. 46). Eukaryotic cells can also develop intracellular concentration gradients of active participants in signaling cascades such as the MAPK pathway (33). Such gradients can also occur in metabolites and even at very small spatial scales, such as the submicrometer intracristae sacs of mitochondria (I. I. Moraru, unpublished data, and Ref. 38). Some recent theoretical work aimed to include diffusion in the mathematical analysis of network control (48), but the applicability of such approaches may be limited to parts of signaling networks that do not span multiple compartments.

How spatial organization controls signal transduction (and even just the role of simple processes such as diffusion) is often difficult to determine without exact knowledge of the morphology and of some of the critical kinetic parameters. Therefore, studies of well-known spatial phenomena (such as intracellular Ca^{2+} waves and transients) have prompted (and required) the development of detailed models with quantitative descriptions of all of the critical components. **FIGURE 3** illustrates an example of the subtleties involved in spatial modeling: Ca^{2+} transients operated by the same signal-transduction mechanism [inositol trisphosphate (IP_3)-induced Ca^{2+} release from the endoplasmic reticulum] show different responses and different mechanisms of regulation in three systems with different morphologies and spatial

scales (L. M. Loew, unpublished data, and Refs. 14, 16, and 61). In large *Xenopus* oocytes, bistability determines a tidal wave, and its propagation is dependent on the propagation of an " IP_3 wave." In differentiated neuroblastoma cells, cell morphology and the uneven distribution of the endoplasmic reticulum shape the characteristics of the transient Ca^{2+} wave, which may be cell-wide, abortive, or limited to the soma, depending on the site of the extracellular stimulus. In Purkinje cell spiny dendrites, the kinetics conspire to provide coincidence detection of repetitive stimuli and generate Ca^{2+} transients that are either localized to a spine or propagating to the dendrite lumen.

The Tools of the Trade: Data, Software, Languages

Finally, we turn to reviewing some of the specific technologies developed over the past few years that were critical for the coming-of-age of computational cell physiology.

Data gathering

There are two fundamental requirements for computational approaches: 1) comprehensive experimental data and 2) a proper analytical framework. The required data are simply concentrations and distributions of molecules (state parameters of the system), their mobility constants, and the binding kinetics and rate laws of their interactions (dynamic parameters of the system). The first type of data is now readily obtainable for gene and protein expression on a large scale (microarrays and mass spectrometry techniques—genome, transcriptome, proteome), or, when necessary, complemented by traditional methods focused on individual molecules.

Much has been made of the uncertainty present in this data, as well as of the sheer lack of quantitative information regarding the dynamic parameters. If history is any guide, however, we believe that technological advances and the combination of directed, large-scale approaches with the work of thousands of individual researchers will make it only a matter of time until the data are available. After all, as little as 15 years ago the (then fledgling) Human Genome Project was argued by some to be intractable, and today we are discussing the future possibility of routine genome sequencing of individual persons for diagnostic purposes (27)! Also, new techniques have been perfected to obtain *in vivo* data with full spatial information. Confocal microscopy can now obtain complete, full, four-dimensional (that is, three-dimensional over time) protein expression data not only in single cells but during the full course of embryogenesis of an entire organism such as *Caenorhabditis elegans* (43) and

can do it so efficiently that a small group of collaborating laboratories could record it for all of the open reading frames of the organism's genome in a reasonable amount of time (W. Mohler, University of Connecticut Health Center, personal communication). Moreover, modeling approaches help direct experiments for acquiring high-quality, quantitative, small-scale data for the critical components of the studied system. Techniques such as fluorescence recovery after photobleaching (FRAP), fluorescence loss in photobleaching (FLIP),

and fluorescence correlation spectroscopy (FCS) today allow accurate measurements of real diffusion coefficients and binding constants in vivo (49).

Data and model exchange

A possibly larger concern is the limited set and relative lack of maturity of available computer tools to analyze and use this data to build models and run simulations (even though there has been an explosion of development efforts in this regard). We can distinguish four broad categories of features of

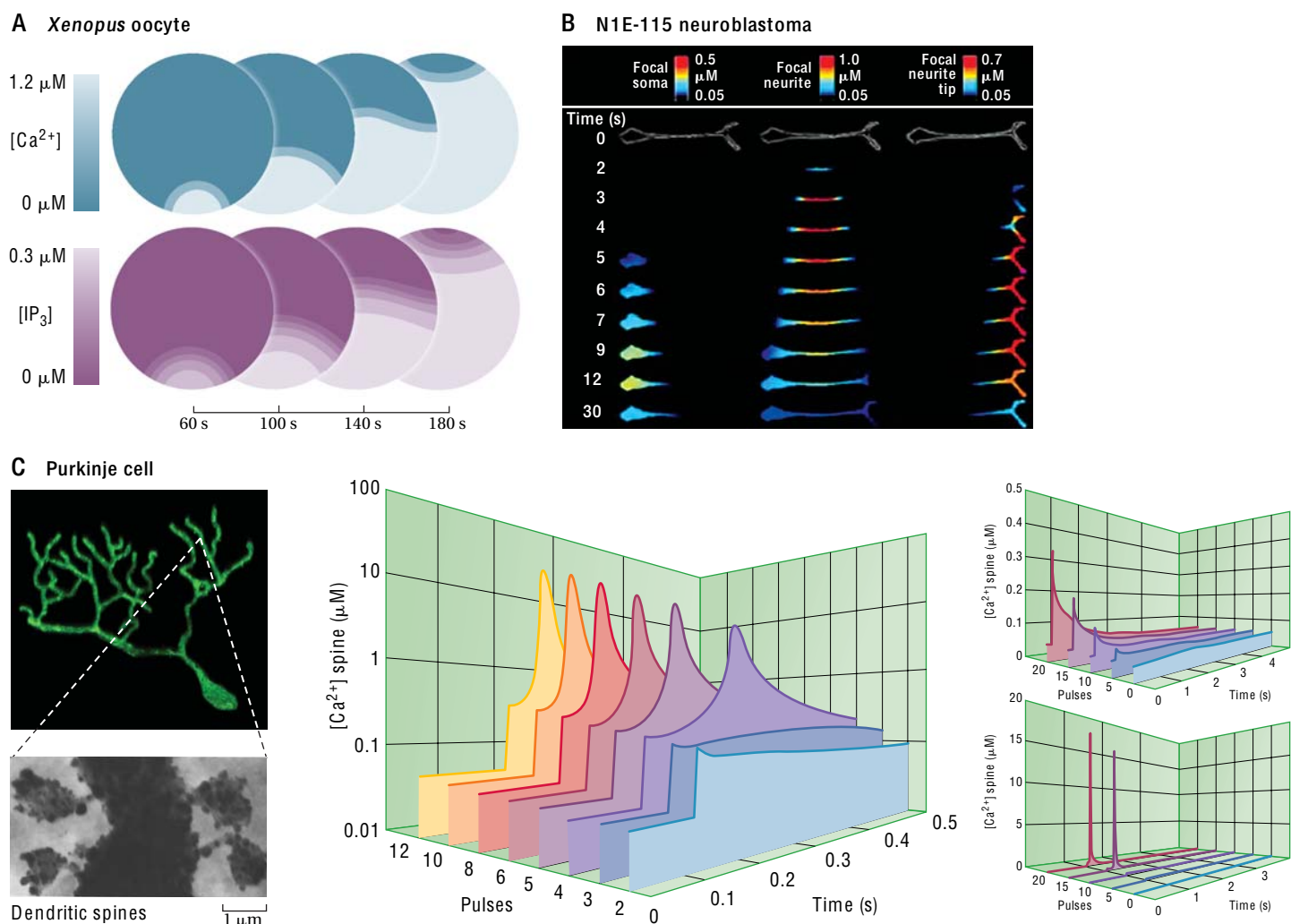


FIGURE 3. Simulations of inositol trisphosphate-dependent Ca^{2+} signaling at different spatial scales

The same basic pathway [inositol trisphosphate (IP_3)-induced Ca^{2+} release] was modeled in cells with widely different morphologies and different parameter values (such as IP_3 receptor density and affinity or Ca^{2+} -buffering capacity). In the giant (millimeter-sized) *Xenopus* oocyte (A), a bistable behavior determines a tidal Ca^{2+} wave after a local trigger of increased Ca^{2+} ($1 \mu\text{M}$) and IP_3 ($0.5 \mu\text{M}$). In differentiated N1E-115 neuroblastoma cells (B), bradykinin induces a transient Ca^{2+} elevation with characteristics dependent on morphology and receptor distribution; a cell-wide Ca^{2+} wave is produced in the case of global stimulation (not shown) or focal stimulation at the neurite (B, middle), but the rise in Ca^{2+} fails to propagate in the case of focal soma stimulation (B, left) or distal neurite stimulation (B, right). Cerebellar Purkinje cells (C, left) have a large dendritic tree with uniform endoplasmic reticulum distribution (visualized by IP_3 receptor fluorescent labeling), and dendrites have thousands of micrometer-size protrusions called spines (shown in an electron micrograph). IP_3 -induced Ca^{2+} rise by stimulation of individual spines has a complex pattern highly dependent on spine morphology and specific kinetic parameters; it may or may not propagate to the dendritic shaft and adjacent spines as a function of the frequency and type of stimulation (not shown) and exhibits supralinearity and thresholding (C, middle, showing the result of increasing number of pulses of IP_3 stimulation) as well as coincidence detection of independent stimuli that produce extracellular Ca^{2+} influx (C, right, showing the result of varying Ca^{2+} stimulations timed at 100 ms or 1 s after a train of pulses of IP_3 stimulation). Note that subsequent signaling after cytosolic Ca^{2+} rise (e.g., Ras/MAPK pathway) at the level of the spine can show significant stochastic effects (see also FIGURE 2). A: adapted from Ref. 61, with permission; B: adapted from *The Journal of Cell Biology*, 1999, 147, 929–935 (16), by copyright permission of The Rockefeller University Press.

these tools: data handling, model building, simulation, and analysis. Many public data repositories and databases exist; at the time of this writing, the Pathway Resource List (<http://www.cbio.mskcc.org/prl/>) had links to 154 internet pathway resources, most of which are databases containing information such as protein-protein interactions or metabolic reactions. However, we need more flexible information management systems to query, annotate, curate, and transform collected data in a format appropriate for modeling, and we are still far from a real biological information system that some researchers have called for (13)—one that could come close to the coverage and functionality of the information management systems that exist in other physical sciences (e.g., GIS, the Geographical Information System).

We have to overcome many problems such as data curation, standards, and even accessibility; for example, the most comprehensive curated database of biological networks at present, Pathways Knowledge Base (Ingenuity, Mountain View, CA) is a commercial product available at only a few major academic institutions. The issue of standards is most critical due to the large variety of public and custom software tools used for modeling and simulation (see below). To facilitate the persistence, comparison, and reuse of computational models among different software platforms, we need a common vocabulary and a common language to describe them. Beyond unique identifiers for the various molecules (which already exist in public databases such as KEGG, SwissProt, GenBank, etc.), we need a comprehensive public ontology that comprises molecules, structures, and biological functions. From the many efforts underway (see Open Biological Ontologies at <http://obo.sourceforge.net> for a current listing), the Gene Ontology and BioCyc projects are quite promising for intracellular networks.

The next level is creating a platform-independent language that would describe the abstractions of a model, replete with quantitative detail, mathematical formulations, and possibly input and output data. Extensible markup language (XML) dialects have been developed in recent years for this purpose, most notably CellML (37) and SBML (17, 28). These two differ both in syntax and in scope: CellML has better support for metadata, spatial information, and mathematical descriptions (using MathML standard), whereas SBML was originally designed to be more specific for pathway and network models. The latter, which was developed as a community effort, has been broadening its scope and seems to be rapidly gaining acceptance as the current standard (currently, more than 60 software packages claim to provide at least some level of SBML support; see <http://www.sbml.org>).

Finally, such ontologies and languages provide the fundamentals for more comprehensive pathway information exchange frameworks such as BioPax (<http://www.biopax.org>), which uses a new standard developed by the W3C consortium, the Web Ontology Language (OWL), that goes beyond XML and Resource Description Framework (RDF, a general standard for, among other uses, metadata representation); it was specifically designed for use by applications that need to not only to present but to process the content of information, thus facilitating machine interpretability.

Model building and simulations

We will skip a detailed review of model inference methods (discussed in THE TOP-DOWN VIEW, above), since most of that type of research uses dedicated software to match the mathematical approaches chosen by the particular group of researchers involved, and we refer the reader to reviews in bioinformatics journals for more technical details.

We will focus now on model building and simulation methods based on a known/hypothesized network connectivity. This was also often achieved as a customized development, but there are an increasing number of “generic” tools that biologists other than the original developers or their collaborators can use (and have used). The fundamental role of these tools is to facilitate a qualitative and quantitative description of the network map (using graphical interfaces or table-based reaction lists), and some will automatically create its mathematical representation (a system of equations). Whether or not created by hand, the latter is the input to simulation engines that use numerical methods to generate data such as the predicted time course of the networks’ state variables given a set of initial conditions. Early entries [e.g., Gepasi (42) and GENESIS/KinetiKit (2)] were focused on simple biochemical reaction networks under conditions in which uniform concentrations of molecules are expected (such as bacterial metabolic reactions). Thus the spatial representation of the system was limited to simple compartmental distributions, which translate in mathematical terms into a system of ordinary differential equations (ODEs). Systems of ODEs have many well-known numerical solvers available, are not very computationally expensive, and allow relatively easy model validation and analysis by methods like flux control analysis and parameter sensitivity calculations. However, they are not sufficient when spatial and stochastic aspects are involved.

When diffusion and actual specific morphologies have to be taken into account, if the system is modeled in a deterministic way, the mathematical representation will include partial differential equations (PDEs). Numerical solution of systems of

PDEs can easily become computationally intractable at physiologically relevant spatial resolution and time courses, and analysis methods such as parameter sensitivity are essentially an open problem. However, since systems of PDEs are a common mathematical formulation for many other processes in physics or engineering (flow, reactors, explosions, etc.), a number of simulation engines for PDEs have been available for quite some time, ranging from relatively simple commercial products, such as FEMLAB (Comsol, Burlington, MA), to high-end versions that can take advantage of massively parallel supercomputers, such as Sandia National Labs MPSalsa (<http://www.cs.sandia.gov/CRF/MPSalsa>) or Pacific Northwest National Labs NWPhys (<http://www.emsl.pnl.gov/nwphys>). Even so, very few tools have been developed that are able to simulate deterministic spatial models and that were specifically designed for modeling biological processes, and most are rather specific (e.g., Continuity from UCSD Cardiac Mechanics Research Group for cardiac modeling). One exception, which is generic and already relatively mature, is The Virtual Cell (<http://www.vcell.org>), developed at the National Center for Cell Analysis and Modeling (36, 56). It is an integrated framework for modeling cell-biological processes that is deployed as a freely accessible distributed application to be used over the Internet, and it includes graphical model-building tools; automated math generation; numerical solvers for simulating both compartmental and spatial problems covering reactions, fluxes, diffusion, advection, electrical potential, and currents on analytic or experimental geometries; and model and results export and analysis tools.

Stochastic simulations have been used for a long time in simulating the dynamics of systems with a small number of states (e.g., modeling the kinetics of membrane channels), usually by using pseudo-random number generation (so-called Monte Carlo simulations) and the Gillespie algorithm (22). Many ad hoc implementations have been recently used to simulate intracellular networks under stochastic regimes, particularly gene-regulatory networks (39). Solving stochastic differential equations is an acceptable approximation when some of the components of the process are abundant enough to be treated using a continuous formulation (i.e., solving the Langevin equation). Stochastic simulations have the additional advantage that they can also solve spatial problems; even though they become computationally expensive when the number of molecules treated discretely is large, they have recently become a realistic alternative to PDE simulations. This is due to a combination of more efficient software implementations, the exponential growth in computer CPU perform-

ance, and the fact that stochastic numerical algorithms lend themselves more easily to parallel processing (thus being able to leverage today's large computer clusters or grid server farms); recent developments in numerical algorithms (8, 21) promise to further expand the range of practical applicability.

MCell (59) is one of the first software platforms that is a general-purpose full stochastic simulator, and although it was initially practical only for small problem domains (e.g., signaling at one neuromuscular junction) and did not allow for direct treatment of binding interactions, the newest version (to be released soon) does not have these limitations (J. Stiles, Pittsburgh Supercomputing Institute, personal communication). Other recently developed stochastic simulators show scalability to the point of metabolic network module simulations, including explicit tracking of all small and large molecules involved (except water and electrolytes) over an entire average-sized bacterial cell (e.g., ChemCell; S. Plimpton, Sandia National Labs, personal communication). Additionally, several large simulation platforms such as E-Cell, Virtual Cell, BioSpice, etc. now include stochastic capabilities.

Conclusion

Proper understanding of the mechanisms of disease as well as effective design and creation of new therapies for disease and aging is hampered by the complexity of cellular interactions. We made a point at the beginning that the essence of understanding is just the ability to predict a system's behavior. Whether this is achieved by finding some (more-or-less simple) sets of rules or laws that govern the system (e.g., "mechanisms") or by simulating a whole network on a computer is ultimately an academic question. Given the complex weave of intracellular interactions governed by many nonlinear and often stochastic relationships, the computer-based approach appears to be the only practical solution.

To elaborate, let us take modern drug development as an example. Traditionally, the bottleneck was the lack of compounds to modulate the activity of a specific target (receptor, enzyme, gene, etc.). Advances in chemistry and high-throughput screening techniques have alleviated this, but even when we have a highly selective and effective modulator for the target of interest, we are often faced with the disappointing result that the final effect on the cell is either not significant or is different from what we expect. Many times this is due to the fact that the target is part of a complex network, which may either be robust with respect to that particular node or have poorly understood emergent proper-

ties. Many of the drugs that were found to be effective in practice by trial and error appear to be quite promiscuous, influencing many cellular targets. Therefore, today the questions most difficult to answer seem to be: 1) which are the best targets to choose and 2) how should they be influenced (inhibit, activate, knockout, overexpress, etc.) to achieve a desired effect? One recent approach has been to identify drug “signatures,” a collection of the effects of a known chemical on many network components (e.g., in the search for kinase inhibitors as potential anticancer drugs), which can be used both to probe the characteristics of the network and to guide in the search for better drugs. Obviously, the ideal answer would be to have a well-characterized model of the pathway(s) of interest through which simulation predictions can identify which is the best combination of “buttons to push” to obtain the desired effect.

There is hope (and fear) that in the not-too-distant future we will be able to reengineer our bodies, to quickly identify and fix whatever goes wrong in most disease states with minimal collateral damage, and even to grow new biological replacement parts that can be as good as, or better than, the (young) original. Whenever that will happen, it will ultimately be due to the development of high-quality tools for analyzing, designing, modeling, and simulating cells and tissues. The components that deal with signaling networks will likely be among the most important, but also most difficult, to develop. ■

References

- Alon U. Biological networks: the tinkerer as an engineer. *Science* 301: 1866–1867, 2003.
- Bhalla US. Use of Kinetikit and GENESIS for modeling signaling pathways. *Methods Enzymol* 345: 3–23, 2002.
- Bhalla US. Signaling in small subcellular volumes. I. Stochastic and diffusion effects on individual pathways. *Biophys J* 87: 733–744, 2004.
- Bhalla US. Signaling in small subcellular volumes. II. Effects on synaptic network properties. *Biophys J* 87: 745–753, 2004.
- Bhalla US and Iyengar R. Emergent properties of networks of biological signaling pathways. *Science* 283: 381–387, 1999.
- Bray D. Reductionism for biochemists—how to survive the protein jungle. *Trends Biochem Sci* 22: 325–326, 1997.
- Bray D. Molecular networks: the top-down view. *Science* 301: 1864–1865, 2003.
- Cao Y, Li H, and Petzold L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J Chem Phys* 121: 4059–4067, 2004.
- Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, and Tyson JJ. Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell* 15: 3841–3862, 2004.
- Csete ME and Doyle JC. Reverse engineering of biological complexity. *Science* 295: 1664–1669, 2002.
- Csete M and Doyle J. Bow ties, metabolism and disease. *Trends Biotechnol* 22: 446–450, 2004.
- Elowitz MB and Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature* 403: 335–338, 2000.
- Endy D and Brent R. Modelling cellular behaviour. *Nature* 409: 391–395, 2001.
- Fall CP, Wagner JM, Loew LM, and Nuccitelli R. Cortically restricted production of IP₃ leads to propagation of the fertilization Ca²⁺ wave along the cell surface in a model of the *Xenopus* egg. *J Theor Biol* 231: 487–496, 2004.
- Fell DA. *Understanding the Control of Metabolism*. London: Portland, 1997.
- Fink CC, Slepchenko B, Moraru II, Schaff J, Watras J, and Loew LM. Morphological control of inositol-1,4,5-trisphosphate-dependent signals. *J Cell Biol* 147: 929–935, 1999.
- Finney A and Hucka M. Systems biology markup language: level 2 and beyond. *Biochem Soc Trans* 31: 6–7, 2003.
- Friedman N. Inferring cellular networks using probabilistic graphical methods. *Science* 303: 799–805, 2004.
- Gardner TS, Cantor CR, and Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339–342, 2000.
- Gardner TS, di Bernardo D, Lorenz D, and Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301: 102–105, 2003.
- Gibson M and Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A* 104: 1876–1889, 2000.
- Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81: 2340–2361, 1977.
- Greenstein JL and Winslow RL. An integrative model of the cardiac ventricular myocyte incorporating local control of Ca²⁺ release. *Biophys J* 83: 2918–2945, 2002.
- Hartwell LH, Hopfield JJ, Leibler S, and Murray AW. From molecular to modular cell biology. *Nature* 402: C47–C52, 1999.
- Hasty J, McMillen D, and Collins JJ. Engineered gene circuits. *Nature* 420: 224–230, 2002.
- Haugh JM, Schooler K, Wells A, Wiley HS, and Lauffenburger DA. Effect of epidermal growth factor internalization on regulation of the phospholipase C- γ 1 signaling pathway. *J Biol Chem* 274: 8958–8965, 1999.
- Hood L, Heath JR, Phelps ME, and Lin B. Systems biology and new technologies enable predictive and preventative medicine. *Science* 306: 640–643, 2004.
- Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, and Wang J; SBML Forum. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19: 524–531, 2003.
- Ideker T, Galitski T, and Hood L. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* 2: 343–372, 2001.
- Ideker T and Lauffenburger D. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol* 21: 255–262, 2003.
- Jordan MI. *Learning in Graphical Models*. Cambridge, MA: MIT Press, 1998.
- Kholodenko BN. Four-dimensional organization of protein kinase signaling cascades: the roles of diffusion, endocytosis and molecular motors. *J Exp Biol* 206: 2073–2082, 2003.
- Kitano H. Systems biology: a brief overview. *Science* 295: 1662–1664, 2002.
- Lander E. Array of hope. *Nat Genet* 21: 3–4, 1999.
- Lazebnik Y. Can a biologist fix a radio? Or, what I learned while studying apoptosis. *Cancer Cell* 2: 179–182, 2002.
- Loew LM and Schaff JC. The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol* 19: 401–406, 2001.
- Lloyd CM, Halstead MD, and Nielsen PF. CellML: its future, present and past. *Prog Biophys Mol Biol* 85: 433–450, 2004.
- Mannella CA, Pfeiffer DR, Bradshaw PC, Moraru II, Slepchenko B, Loew LM, Hsieh C, Buttle K, and Marko M.

- Topology of the mitochondrial inner membrane: dynamics and bioenergetic implications. *IUBMB Life* 52: 93–100, 2001.
39. McAdams HH and Arkin A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* 15: 65–69, 1999.
 40. McAdams HH and Shapiro L. A bacterial cell-cycle regulatory network operating in time and space. *Science* 301: 1874–1877, 2003.
 41. Mellman I and Misteli T. Computational cell biology. *J Cell Biol* 161: 463–464, 2003.
 42. Mendes P. Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem Sci* 22: 361–363, 1997.
 43. Mohler WA and White JG. Multiphoton imaging of embryonic development. In: *Imaging Neurons: A Laboratory Manual*, edited by Yuste R, Lanni F, and Konnerth A. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 2000, sect. 3, chapt. 21, p. 21.11–21.21.
 44. Nachman I, Regev A, and Friedman N. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics* 20: i248–i256, 2004.
 45. Neves SR and Iyengar R. Modeling of signaling networks. *Bioessays* 24: 1110–1117, 2002.
 46. Noble D. Modeling the heart. *Physiology* 19: 191–197, 2004.
 47. Novak B and Tyson JJ. Numerical analysis of a comprehensive model of M-phase control in *Xenopus* oocyte extracts and intact embryos. *J Cell Sci* 106: 1153–1168, 1993.
 48. Peletier MA, Westerhoff HV, and Kholodenko BN. Control of spatially heterogeneous and time-varying cellular reaction networks: a new summation law. *J Theor Biol* 225: 477–487, 2003.
 49. Phair RD and Misteli T. Kinetic modeling approaches to in vivo imaging. *Nat Rev Mol Cell Biol* 2: 898–907, 2001.
 50. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, and Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551–1555, 2002.
 51. Resat H, Ewald JA, Dixon DA, and Wiley HS. An integrated model of epidermal growth factor receptor trafficking and signal transduction. *Biophys J* 85: 730–743, 2003.
 52. Rice JJ, Jafri S, and Winslow RL. Modeling gain and gradedness of Ca^{2+} release in the functional unit of the cardiac diadic space. *Biophys J* 77: 1871–1884, 1999.
 53. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, and Friedman N. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34: 166–176, 2003.
 54. Selinger DW, Wright MA, and Church GM. On the complete determination of biological systems. *Trends Biotechnol* 21: 251–254, 2003.
 55. Slepchenko BM, Schaff JC, Carson JH, and Loew LM. Computational cell biology: spatiotemporal simulation of cellular events. *Annu Rev Biophys Biomol Struct* 31: 423–441, 2002.
 56. Slepchenko BM, Schaff JC, Macara I, and Loew LM. Quantitative cell biology with the Virtual Cell. *Trends Cell Biol* 13: 570–576, 2003.
 57. Sourjik V. Receptor clustering and signal processing in *E. coli* chemotaxis. *Trends Microbiol* 12: 569–576, 2004.
 58. Stelling J, Sauer U, Szallasi Z, Doyle FJ, and Doyle J. Robustness of cellular functions. *Cell* 118: 675–685, 2004.
 59. Stiles JR and Bartol TM. Monte Carlo methods for simulating realistic synaptic microphysiology using MCell. In: *Computational Neuroscience: Realistic Modeling for Experimentalists*, edited by De Schutter E. Boca Raton, FL: CRC, 2001, p. 87–127.
 60. Tyson JJ, Chen KC, and Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 15: 221–231, 2003.
 61. Wagner J, Fall CP, Hong F, Sims CE, Allbritton NL, Fontanilla RA, Moraru II, Loew LM, and Nuccitelli R. A wave of IP_3 production accompanies the fertilization Ca^{2+} wave in the egg of the frog, *Xenopus laevis*: theoretical and experimental support. *Cell Calcium* 35: 433–447, 2004.
 62. Wiley HS, Shvartsman SY, and Lauffenburger DA. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* 13: 43–50, 2003.
 63. Wolf DM and Arkin AP. Motifs, modules and games in bacteria. *Curr Opin Microbiol* 6: 125–134, 2003.