The Virtual Cell: a software environment for computational cell biology

Leslie M. Loew and James C. Schaff

The newly emerging field of computational cell biology requires software tools that address the needs of a broad community of scientists. Cell biological processes are controlled by an interacting set of biochemical and electrophysiological events that are distributed within complex cellular structures. Computational modeling is familiar to researchers in fields such as molecular structure, neurobiology and metabolic pathway engineering, and is rapidly emerging in the area of gene expression. Although some of these established modeling approaches can be adapted to address problems of interest to cell biologists, relatively few software development efforts have been directed at the field as a whole. The Virtual Cell is a computational environment designed for cell biologists as well as for mathematical biologists and bioengineers. It serves to aid the construction of cell biological models and the generation of simulations from them. The system enables the formulation of both compartmental and spatial models, the latter with either idealized or experimentally derived geometries of one, two or three dimensions.

SIMULATION

The information explosion in biology provides unprecedented opportunities for the biomedical research community. However, to fully seize these opportunities we need new tools that can help us analyze this immense amount of data. Clearly, database technology applied to genomic, proteomic and biochemical pathway information – the new discipline of bioinformatics – can help us to extract the data that might be relevant to a complex biological process and begin to discern patterns and relationships. But the bioinformatics arsenal also requires tools that enable researchers to test whether a set of interacting molecules and structures can produce an observed behavior. With respect to cell biology, such a tool should allow the construction of quantitative models of processes as diverse as signal transduction, nuclear transport and mitochondrial respiration.

The 'Virtual Cell' is a unique computational tool that is being developed at the University of Connecticut Health Center (Farmington, CT, USA) to help address this need[1–4]. It is a general software system that allows biologists with little training in physics and mathematics to engage in computational cell biology. A biology-oriented graphical user interface provides for the assembly of models by specifying the molecules, reactions and structures involved. By providing a structured environment for pulling together related pieces of quantitative data, it enables the construction of complex spatial models of biological processes. From these models, simulations are then produced by the software and the predictions of these simulations can be directly compared with experimental data. If the simulation does not match the experiment, the model must be an incomplete or faulty description of the cell biology and must be modified. If the simulations are consistent with the experiment, new simulations with different conditions can be used to predict the results of new experiments that can further test the limits of the applicability of the model. This procedure is simply a restatement of the classical scientific method for the case of a complex set of cell biological hypotheses – a cell biological model.

The Virtual Cell system is also designed to be useful for experienced modelers such as bioengineers and mathematical biologists. It allows the direct entry of mathematical equations that describe a model using a declarative language (Virtual Cell Mathematics Description Language, VCMDL). The mathematics is then automatically translated into C++ programming code, which can than be sent to the NUMERICS SOLVERS (see glossary). Thus, modelers are relieved of the drudgery of writing *ad hoc* code for

Leslie M. Loew*
James C. Schaff
Center for Biomedical Imaging Technology, Dept of Physiology, University of Connecticut Health Center, Connecticut, Farmington, CT 06030, USA.
*e-mail: les@volt.uchc.edu

Glossary

Local sensitivity analysis: Determination of the relative changes over time in the value of model variables for a differential change in the value of a parameter.
Mesh: A grid of spatial elements used in numerical solutions to partial differential equations – the finer the mesh, the more accurate the solution.
Numerics solvers: Software modules that implement a numerical algorithm to solve a class of mathematical equations.
Ordinary differential equation (ODE) simulation: A simulation of a model that requires only the solution of ordinary differential equations corresponding to the rates of a system of biochemical reactions. Such models do not explicitly treat cell geometry or diffusion.
Pseudo-steady-state approximations: If a reaction is fast on the time-scale of the overall process that is being modeled, it can be treated as if it is in equilibrium at each time point in the simulation. This avoids solving the differential equations for these variables and can dramatically improve the speed of calculations.
Spatial partial differential equation (PDE) simulation: A simulation of a model that requires the solution of partial differential equations corresponding to the rates of a system of biochemical reactions plus diffusion of molecules within a specified cellular geometry.
XML: Extensible Markup Language. XML is the standard specification for producing structured text-based descriptions of data that enables the interchange of information between different software applications (http://www.xml.org/xml/xmlfaq2.shtml).

 0167-7799/01/$ – see front matter PII: S0167-7799(01)01740-1

Box 1. Websites of relevant software tools

- Virtual Cell: http://www.nrcam.uchc.edu
- Gepasi: http://www.gepasi.org
- Jarnac/Scamp: http://members.tripod.co.uk/sauro/biotech.htm
- DBSolve: http://websites.ntl.com/~igor.goryanin
- E-Cell: http://www.e-cell.org
- BioSpice: http://www.lbl.gov/~aparkin
- StochSim: http://www.zoo.cam.ac.uk/comp-cell/StochSim.html
- MCell: http://www.mcell.psc.edu
- Genesis/Kinetikit: http://www.ncbs.res.in/~bhalla/kkit/index.html
- SBML: http://www.cds.caltech.edu/erato
- CellML: http://www.cellml.org

every new modeling task. Furthermore, a VCMDL description of a model can be produced directly and automatically from a model that has been created within the biological interface. Thus, this dual interface makes the Virtual Cell environment highly effective for promoting and supporting interactions between experimental biologists and mathematical modelers and will serve to bridge the cultural gap between these traditionally separate communities.

Several other important software tools are being developed that might be applied to cell biology and have strengths that nicely complement those of the Virtual Cell system, these include: (1) GEPASI (Ref. 5), Jarnac/Scamp (Ref. 6) and DBSolve (Ref. 7) are simulators for complex metabolic and chemical pathways and include extensive analytical and optimization tools; (2) ECELL (Ref. 8) is a computational system for constructing whole cell models and a model of a self-sustaining primitive cell has been completed; (3) BioSpice (Ref. 9) is a general purpose modeling system with particular strengths in analyzing prokaryotic genetic circuits; (4) StochSim (Ref. 10) permits the user to follow individual molecules in a regulatory pathway through stochastic simulations; and (5) MCell (Ref. 11) is another stochastic simulator that is designed for the study of subcellular processes such as synaptic transmission and is the only system other than Virtual Cell that can explicitly accommodate structural information. Also noteworthy is an adaptation for signal transduction pathways (Kinetikit) of the popular neuroscience modeling software GENESIS (Ref. 12). Although we have provided references for all of these software tools, the best information about the availability and current status of these packages is found on the developers' websites (Box 1). Also, it is doubtful that this list is exhaustive; it is clear that the collection of tools for computational approaches to cell biology is growing rapidly.

A look at the Virtual Cell system

The modeling process within the Virtual Cell framework can be understood with reference to Figure 1. Experiment provides the motivation for the development of a model as well as the data that are used to construct the model. The data includes the identity of the molecules involved, their reactions and transport properties, where they are compartmentalized within the cell, and the topological organization of those compartments. This comprises the physiology of the cellular process that is being modeled. LOCAL SENSITIVITY ANALYSIS tools are available and links to external databases are planned (by mid-2002) to aid in the choice of parameters and model components. For spatial simulations, the various compartments have to be mapped to the appropriate geometries. These can be idealized analytical geometries, fully resolved structures derived from digital microscope images, or continuously distributed compartments within resolved structures. This scheme allows the same physiology to be reused with many geometries and facilitates ready adaptation and modification of models. Once the physiology is mapped to the geometry and the initial and boundary conditions are specified, the model is fully defined. The framework automatically converts the biological mechanisms to a corresponding mathematical system that incorporates mass conservation relationships and PSEUDO-STEADY-STATE APPROXIMATIONS. The mathematical description, expressed in VCMDL, can be further edited and refined at this point. When the appropriate MESH sizes and time steps for the simulation have been chosen, the model is sent to the appropriate solver and a simulation is generated. A comparison of the simulation results with the original experimental data dictates whether the model will need to be modified to account for what has been observed experimentally. Data visualization resources are provided for navigating the huge simulation datasets. These include selectable data dumps in spreadsheet format and pseudo-colored images coding the magnitude of any variable at any point in time as well as the corresponding Quicktime movies. Because the simulations produce the same types of spatial and temporal records that can be obtained by experiment, the predictions of a model can be analyzed using the same statistical and/or image analysis methods used to analyze experiments.

An example screen from the user interface is shown in Fig. 2 and illustrates the key operation of linking physiology to geometry in a model. The left part of the screen shows the topology that had been entered during a previous operation. It simply identifies the cytosol as being enclosed within the extracellular space and the endoplasmic reticulum as being within the cytosol. All the molecules and reactions associated with these compartments and with the membranes separating them are also specified in earlier input operations. The right hand part of the screen shows an image of a neuronal cell that had been segmented into an extracellular and a cytoplasmic region during a previous operation. The linkage between the physiology and the geometry is achieved by simply dragging connecting lines

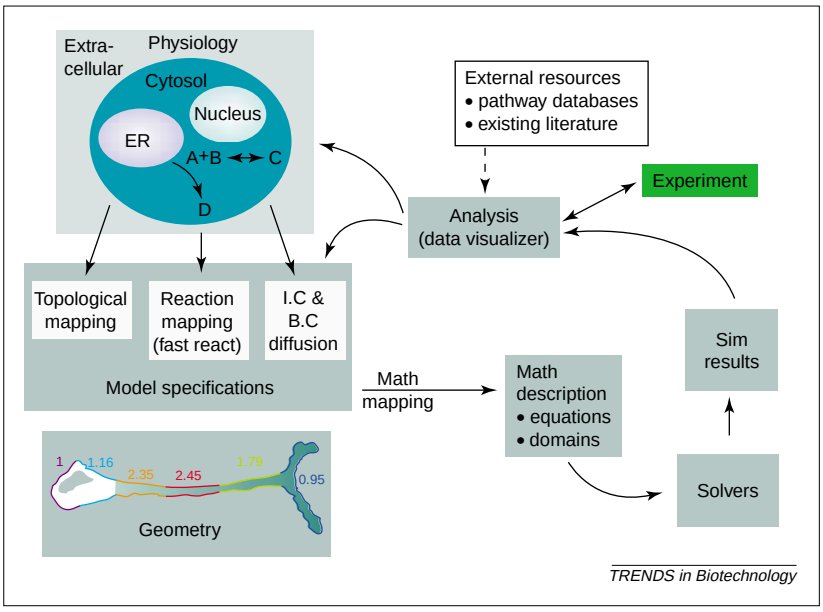

Fig. 1. The design of the Virtual Cell imposes a strict interplay between model development and experiment during the modeling process. Inputs to the model can be derived from the user's own experiments as well as the literature (plans to link the system to external databases are being developed). The 'physiology' includes the topological arrangements of compartments and membranes, the molecules associated with each of these, and the reactions between the molecules. The separately defined 'geometry' is a spatial description of the layout of the compartments in 0 (i.e. non-spatial), one, two or three dimensions; it can be derived from either analytical expressions or from an experimental image acquired from a microscope. The numbers represent the relative surface densities of the BKR. Linking the topology in the physiological description to an appropriate geometry specifies the model; initial conditions (I.C.) and boundary conditions (B.C.) must also be entered to fully define a mathematical description of the model. The mathematical description is then automatically generated and the simulation is run via an appropriate solver. Simulation results can be displayed in a variety of formats and can also be exported as images, movies or spreadsheet data. The results of these simulations are then compared with experiments, often in identical formats, to determine how well the model predicts the behavior of the overall system. Based on these comparisons, missing or incorrect features of the model can be identified.

between the corresponding regions on the left and right. It is noteworthy that the endoplasmic reticulum is unresolved in the cell image, which originated from an optical microscope. By connecting both the cytosol and endoplasmic reticulum compartments in the topology to the same 'cytosol' region of the segmented image, the endoplasmic reticulum is automatically assigned to be continuously distributed within the cytosol. In the lower portion of the screen, the surface-to-volume ratio for the endoplasmic reticulum and its volume fraction within its parent compartment (i.e. the cytosol) are specified. Indeed, the physiological model needn't be assigned to a geometry at all. Simulations from such compartmental models require much less computational time than spatial models and can be run during preliminary optimizations of model behavior before running full spatial models.

An object-oriented distributed software architecture has been used for the Virtual Cell. Several standard solvers including solvers for stiff equations and variable time-step solvers are available for the solution of ordinary differential equations (ODEs). The finite volume method[13,14] is currently used for partial differential equation (PDE) problems. Both deterministic and stochastic physical formulations have been implemented in the C++ library. However, only deterministic models can be fully developed using the current interface. Physiologies, geometries, mathematical descriptions and simulation results are stored in a central database that maintains the privacy of user models while providing a mechanism for model sharing. In addition to allowing the maintenance of these database resources, the client-server architecture has the advantage of allowing frequent updates to the software and rapid responses to problems uncovered by users. It also permits the use of a high performance cluster computer system at our site for the numerics service. Technology based on the JAVA programming language allows access to the software via a web browser. Access to the Virtual Cell is available on the National Resource for Cell Analysis and Modeling website (http://www.nrcam.uchc.edu).

An example of a Virtual Cell application: calcium dynamics in a neuronal cell

Inositol-1,4,5-triphosphate (IP_3)-mediated calcium signaling is found in many cells in response to many external stimuli[15–17]. Figure 3 shows the pathway for this overall process. Fluorescent indicator dyes have allowed the calcium release following IP_3 generation to be studied in great detail. However, the absence of an indicator dye for IP_3 itself has proven a hindrance in identifying the spatial and temporal characteristics of IP_3 generation, propagation, and degradation during a cell-signaling event. In addition, there is a great deal of morphological variation between cell types. In particular, neurons have complex geometries, comprising axons and dendritic trees that have much higher surface-to-volume ratios than the soma. How these morphologic features control the patterns of IP_3 signaling has not yet been determined.

In differentiated neuroblastoma cells, for example, bradykinin (BK) triggers IP_3-dependent calcium waves that consistently start in the neurite proximal to the soma and rapidly propagate in both directions. Using calcium imaging, quantitative uncaging of microinjected IP_3 and simulations from the Virtual Cell, we found that IP_3 levels build up in the neurite at a rate and to an extent much greater than in the soma[18,19]. Simulations and experiments using focal applications of BK confirmed that the proximal segment of the neurite is the crucial region for a response to a BK stimulus and is necessary and sufficient to initiate and propagate the calcium signal to other regions of the cell. The interplay of a high density of calcium stores in the soma and a rapid rise of $[IP_3]_{cyt}$ in the neurite can explain these results. The latter is attributed to a high surface-to-volume ratio in the neurite and might reflect a general principle whereby the neuronal morphology amplifies signals from diffusible second messengers in axons and dendrites.

This study represents a good example of the interplay of experiment and modeling. The first

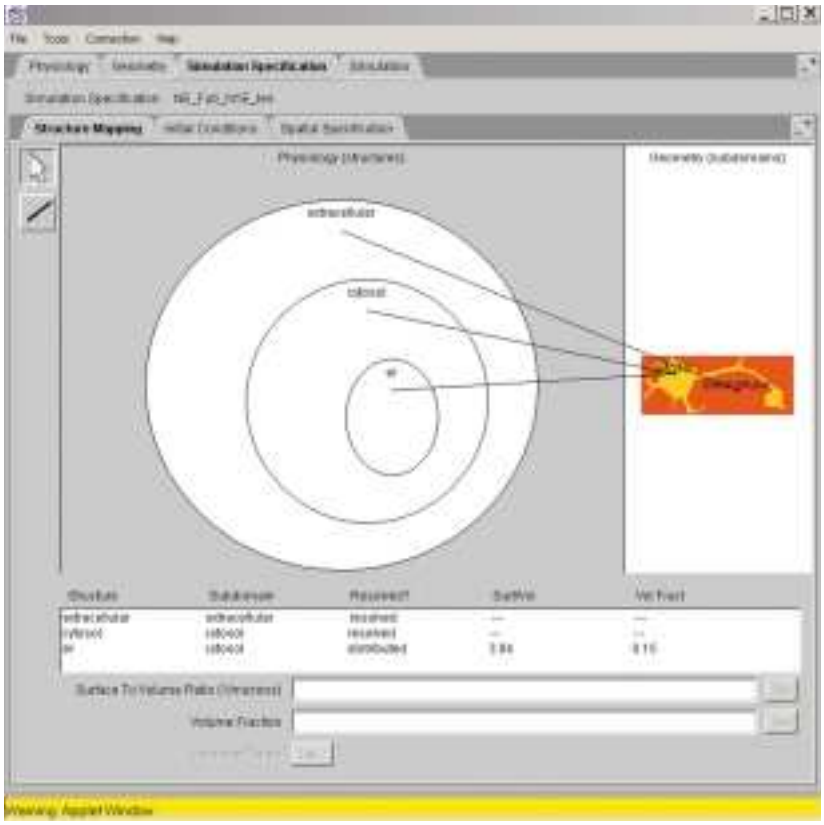

Fig. 2. The graphical user interface of the Virtual Cell appears as a JAVA applet within a web browser. This example screen illustrates how cell topology is linked to a specific experimental geometry.

column of Fig. 4 shows the results of an experiment for the response of intracellular Ca^{2+} to BK. These data were collected on a fast digital-imaging microscope using the fluorescence from the indicator fura-2 to record the spatiotemporal changes in $[Ca^{2+}]_{cyt}$. The initial calcium increase is observed in the middle of the neurite after 2.2 seconds, spreading bi-directionally to the soma and growth cone and reaching a peak $[Ca^{2+}]_{cyt}$ everywhere of ~1 μM. This pattern was observed in all the cells studied as long as they had this characteristic neuronal morphology. A quantitative model of the process shown in Fig. 3 was constructed using the Virtual Cell software to determine how all these individual components could interact to produce the observed calcium wave. We obtained the data for the model both from the literature and from experiments in our own laboratory. Indeed, in several instances previously reported experiments had to be repeated when it was found that a model based on the literature data could not predict the observed calcium dynamics. One of the most interesting findings that came out of this iterative modeling and experiment process was the prediction that the observed calcium wave could only be reproduced if the density of the calcium store in the soma was approximately twice as high as that in the neurite. Subsequently, the results of immunofluorescence studies showed that the endoplasmic reticulum had precisely this predicted non-uniform distribution. The simulation for $[Ca^{2+}]_{cyt}$ resulting from the completed model can be directly compared with the experiment in the first two columns of Figure 4. This illustrates how simulations can be mapped to the same geometry as the experiment; thus, the investigator can analyze simulation results with the same familiar tools that are used to analyze and reduce experimental image data.

There is no available fluorescent indicator for $[IP_3]$, therefore the model provides a unique view of the spatial and temporal distribution of this key metabolite. The spatiotemporal pattern of $[IP_3]_{cyt}$ predicted by the model is shown in the third column of Figure 4. The calculated $[IP_3]_{cyt}$ dynamics show a rapid buildup in the neurite to a peak of ~10 μM, whereas it increases more slowly and to lower peak concentrations (~3 μM) in the soma. The production of IP_3 is much faster than its diffusion throughout the intracellular volume and it also outpaces the rate of degradation through putative cytosolic kinase-based and phosphatase-based pathways. Therefore, because IP_3 is produced from the plasma membrane, the cytosolic concentrations of IP_3 will rise faster and with greater maximum amplitude in the neurite than in the soma. This is primarily because of the high surface-to-volume ratio of the neurite compared with the soma. Interestingly, despite the higher level of IP_3 in the neurite, the maximum calcium amplitudes are similar in all parts of the cell. This is because the non-uniform distribution of calcium stores balances the non-uniform levels of IP_3. (The reader is referred to the original papers[18,19] for additional experiments that were suggested by predictions of the model and full details of the model components; movies of both the experimental and simulation results can be found at http://www.nrcam.uchc.edu/.)

Future prospects

The current version of the Virtual Cell system can handle a large range of modeling problems encompassing reaction–diffusion processes in arbitrary geometries. However, problems that require a changing geometry, such as cell migration or mitosis, will require the system to be significantly enhanced. Also, the system is currently able to treat only some types of stochastic processes such as Brownian motion, directed particle motility along microtubule tracks, and the reaction of individual particles with continuously distributed molecules (this work is unpublished, but a simple example is available on our website: http://www.nrcam.uchc.edu/rna_traffick_dir/rna_traffick.html). For situations in which the number of interacting molecules is too low for a continuous description, there is a need to expand the stochastic formulations to include the treatment of reactions between discrete molecular species. We also need to develop a discrete state treatment for models of single ion-channel currents and locations. Adaptations of the system architecture and the user interface to fully accommodate stochastic models are also planned. In addition, we hope to incorporate more efficient solvers within the numerics service including algorithms that take advantage of massively parallel computer architectures. This, along with external database support, will enable the

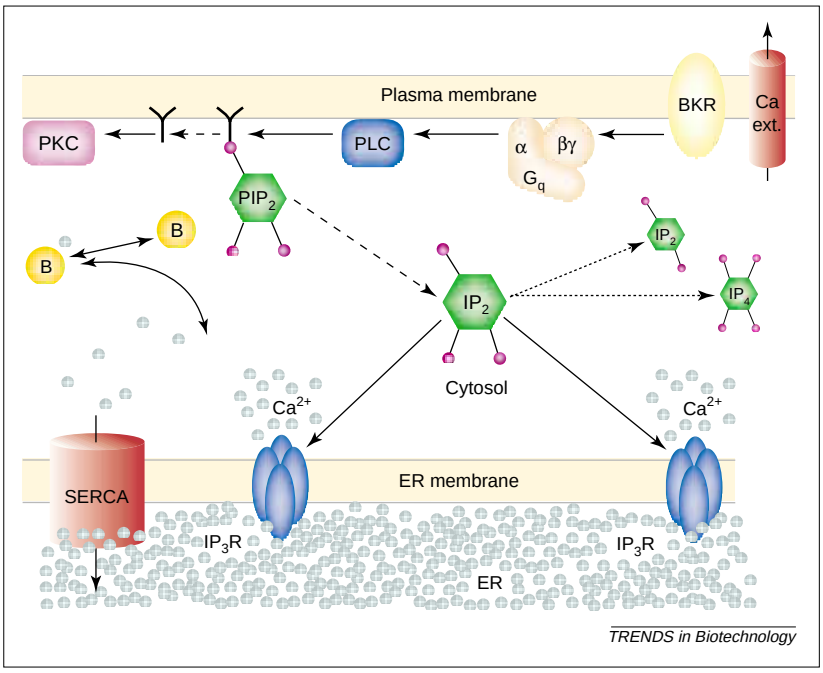

Fig. 3. The pathway for bradykinin-induced calcium release in differentiated neuroblastoma cells. The cascade of events is triggered when bradykinin (BK) binds to its receptor (BKR) in the plasma membrane. This sets off a G-protein cascade (the G-protein complex is shown as αGq, βγ) that activates phospholipase C (PLC) that, in turn, hydrolyzes the glycerolphosphate bond in phosphatidylinositolbisphosphate (PIP_2). One product of this hydrolysis goes on to activate protein kinase C (PKC), but more importantly for our model it releases IP_3 from the membrane. The IP_3 is free to diffuse through the cytosol, be degraded by phosphatases and kinases, and bind to its receptor (IP_3R) in the endoplasmic reticulum (ER) membrane. The IP_3R is a calcium channel that is triggered to open when IP_3 is bound and when calcium itself binds to an activation site; a slower inactivation process also pertains to this channel. The calcium that is thus released binds to calcium buffers (B) in the cytosol including the fluorescent calcium indicator that is used experimentally to visualize the $[Ca^{2+}]_{cyt}$. Finally, calcium is pumped back into the ER via a calcium ATPase (SERCA) and can also be extruded from the cell by several exchange and pump mechanisms (Ca ext.).

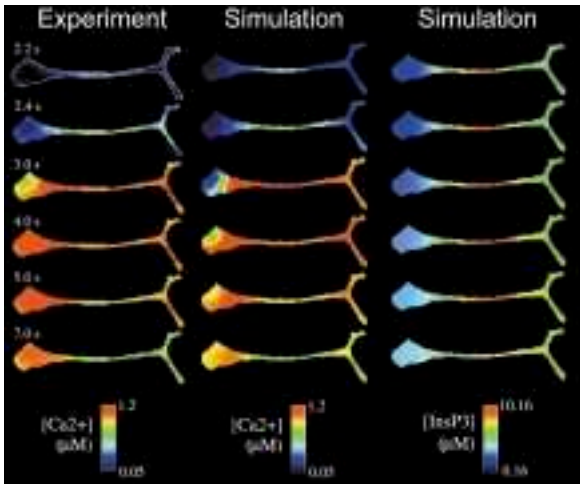

Fig. 4. The outputs of the Virtual Cell allow direct comparison of experiment with simulation. The left column shows the experimental calcium changes following addition of BK at time 0 s in a differentiated N1E-115 neuroblastoma cell as determined by the fluorescent calcium indicator fura-2. The center column displays the corresponding output of the Virtual Cell simulation. The right column displays the output of the simulation for $[IP_3]$, illustrating that the simulation permits estimation of the spatiotemporal distribution of molecules that are not accessible experimentally. The scale at the bottom of each column defines the range of values coded by the colors in the images. (For additional details, see Ref. 19.)

construction of larger scale biochemical network models within a high-resolution cellular geometry.

Another important future challenge – and opportunity – will be the development of mechanisms for the interchange of models between different modeling tools. For example, the sophisticated analytical capabilities of large-scale modeling systems such as GEPASI could be used to optimize and reduce a model at the ODE level before it is migrated to the Virtual Cell for a SPATIAL PDE SIMULATION. There has been a general interest in developing model interchange standards using XML. The Virtual Cell development team is currently participating in two such efforts. CellML (http://www.cellml.org) is a markup language being developed by Physiome Sciences Inc. (Princeton, NJ, USA) in collaboration with the Bioengineering Department at the University of Auckland (Auckland, New Zealand). SBML (http://www.cds.caltech.edu/erato/) is being developed by a team at California Institute of Technology (Pasadena, CA, USA) in collaboration with the ERATO-Kitano Symbiotic Systems Project in Japan (Tokyo, Japan). Although these two efforts are not identical, they both aim to allow scientists to share models across model-building software platforms. This will also facilitate the storage and retrieval of model components as an aid to the construction of new models. The cooperative spirit that underlies these efforts bodes well for the future of the field of computational cell biology.

It is important to emphasize that the Virtual Cell is primarily designed for the cell biology community in that it enables the construction of spatial models in a logical and user-friendly manner and can therefore be used in association with quantitative live cell microscopy. Together with other emerging software tools, it will have a major impact on the budding field of computational cell biology. The emergence of computational cell biology can be compared with molecular modeling, which uses physical principles to calculate protein and nucleic acid structure and dynamics. As recently as ten years ago, molecular modeling was the province of specialists in quantum and statistical mechanics with access to (what were then) super computers. It is only with the increased power of computers and the development of sophisticated yet user-friendly software that molecular modeling has now become a standard tool in drug development and structural biology research. Today there are at least a dozen companies producing molecular modeling software that is widely used both in academia and in the pharmaceutical industry for drug design. Cell biological modeling will appeal to an even broader research community because it will be applicable to a broader range of biological problems. Furthermore, it promises to be an important educational tool. We believe that the Virtual Cell system will be an important medium for establishing computational approaches within the community of experimental cell biologists.

Acknowledgements
The authors are pleased to acknowledge the contributions of Boris Slepchenko, Ion Moraru, Charles Fink, James Watras, Yung-sze Choi, John Wagner, and Diana Resasco. This work was supported through a Biomedical Technology Resource Grant (RR13186) from the National Institutes of Health National Center for Research Resources.

References

1 Schaff, J. *et al.* (1997) A general computational framework for modeling cellular structure and function. *Biophys. J.* 73, 1135–1146
2 Schaff, J. and Loew, L.M. (1999) The Virtual Cell. In *Biocomputing: Proceedings of the 1999 Pacific Symposium.* (Altman, R.B., Dunker, A.K., Hunter, L., Klein, T.E., Lauderdale, K., eds) pp. 228–239, World Scientific Publishing Co.
3 Schaff, J.C. *et al.* (2000) Physiological modeling with the Virtual Cell framework. In *Methods in Enzymology* (Vol. 321) (Johnson, M., ed.) pp. 1–23, Academic Press
4 Schaff, J.C. *et al.* (2001) Analysis of non-linear dynamics on arbitrary geometries with the Virtual Cell. *Chaos* 11, 115–131
5 Mendes, P. and Kell, D. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 14, 869–883
6 Sauro, H.M. (1993) SCAMP: a general-purpose simulator and metabolic control analysis program. *Comput. Appl. Biosci.* 9, 441–450
7 Goryanin, I. *et al.* (1999) Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics* 15, 749–758
8 Tomita, M. *et al.* (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15, 72–84
9 McAdams, H.H. and Arkin, A. (1998) Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* 27, 199–224
10 Morton-Firth, C.J. *et al.* (1999) A free-energy-based stochastic simulation of the Tar receptor complex. *J. Mol. Biol.* 286, 1059–1074
11 Stiles, J.R. *et al.* (1998) Monte Carlo simulation of neurotransmitter release using MCell, a general simulator of cellular physiological processes. In *Computational Neuroscience.* (Bower, J., ed.) pp. 279–284, Plenum Press
12 Bhalla, U.S. and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways *Science* 283, 381–387
13 Patankar, S.V. (1980) *Numerical Heat Transfer and Fluid Flow.* Taylor and Francis
14 Smith, G.D. (1985) *Numerical Solution of Partial Differential Equations: Finite Difference Methods.* Clarendon Press
15 Berridge, M.J. (1993) Inositol trisphosphate and calcium signaling. *Nature* 361, 315–325
16 Putney, J.W. Jr and Bird, G.S. (1993) The inositol phosphate-calcium signaling system in nonexcitable cells. *Endocrinol. Rev.* 14, 610–631
17 Berridge, M.J. (1998) Neuronal calcium signaling. *Neuron* 21, 13–26
18 Fink, C.C. *et al.* (1999) Morphological control of inositol-1,4,5-trisphosphate-dependent signals. *J. Cell Biol.* 147, 929–935
19 Fink, C.C. *et al.* (2000) An image-based model of calcium waves in differentiated neuroblastoma cells. *Biophys. J.* 79, 163–183

RNase P: from biological function to biotechnological applications

César Cobaleda and Isidro Sánchez-García

The M1 RNA subunit of *Escherichia coli* RNase P is a ribozyme responsible for the catalytic activity of the complex. It removes the 5′ leader sequence from tRNA precursors to form mature tRNAs. M1 recognizes its target mainly on the basis of its structure and this allows the design of modified ribozymes engineered to destroy other molecules without the need for special sequences in the targeted mRNAs. M1 is thus an ideal tool to eliminate the tumourigenic chimeric messengers created after chromosomal translocations. These results have direct implications for cancer therapeutics and molecular biology in general.

César Cobaleda*
Isidro Sánchez-García*
Instituto de Biología Molecular y Celular del Cáncer (IBMCC), Centro de Investigación del Cáncer, CSIC/Universidad de Salamanca, Campus Unamuno, 37007-Salamanca, Spain.
*e-mails: ccobalz@usal.es and isg@usal.es

The biological function of RNase P is to catalyse the hydrolytic reaction that removes the 5′ leader sequence of pre-tRNAs and other small RNAs (Ref. 1) (Fig. 1). It is a ribonucleotide protein and has been found in all cells, bacterial, archaebacterial and eukaryotic. In *Escherichia coli*, its protein subunit (C5) is not essential for the enzymatic activity of the complex. This catalytic activity resides in the RNA component, the M1 RNA ribozyme[2].

The interaction of M1 RNA with its natural substrates depends on the recognition of substrate structure rather than of some specific sequences or of nucleotides in the target. This is in contrast to other kinds of ribozymes, such as the widely used hammerhead or hairpin-based ones[3]. The structural elements that M1 recognizes in its targets are[4,5] (Fig. 1): (1) a double-stranded RNA region (corresponding to the aminoacyl or T-stem of pre-tRNA); and (2) a 3′ CCA unpaired sequence. Moreover, a 5′ leader sequence is required as a string of single-stranded RNA that is a part of substrate (in the strand that has to be cut). Nevertheless, a distinction must be made when the M1 RNA reaction is in the presence of its cognate C5 protein: the RNA structure primarily recognized by the M1 RNA–C5 protein complex includes the acceptor stem and T stem of the pre-tRNA, less the T loop, and not the 3′ CCA.

Therefore, if these structural requirements can be provided, any given sequence can in principle be chosen as M1 RNA target. To achieve this, a so-called ‘guide sequence’ (GS) has to be designed to hybridize with the target mRNA. This GS will mimic the

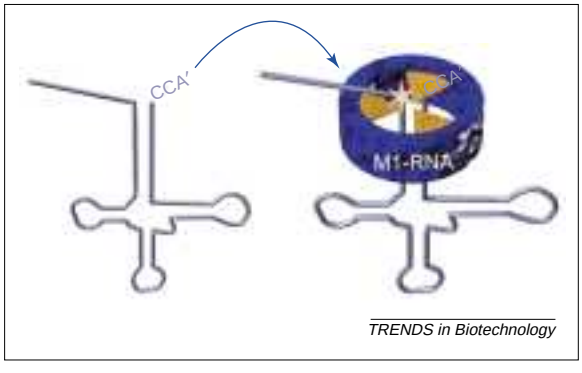

Fig. 1. Biological function of M1 RNA from *Escherichia coli* RNase P. Pre-tRNA (left) is the natural target of RNase P, which recognizes the double-stranded RNA structure of the pre-tRNA T-stem and the 3′ CCA unpaired stretch, and cleaves the 5′ leader to release the mature tRNA.

 0167-7799/01/$ – see front matter PII: S0167-7799(01)01741-3