Compartmental and Spatial Rule-Based Modeling with Virtual Cell

Michael L. Blinov,1,* James C. Schaff,1 Dan Vasilescu,1 Ion I. Moraru,1 Judy E. Bloom,1 and Leslie M. Loew1,*
1R. D. Berlin Center for Cell Analysis and Modeling, University of Connecticut School of Medicine, Farmington, Connecticut

ABSTRACT In rule-based modeling, molecular interactions are systematically specified in the form of reaction rules that serve as generators of reactions. This provides a way to account for all the potential molecular complexes and interactions among multivalent or multistate molecules. Recently, we introduced rule-based modeling into the Virtual Cell (VCell) modeling framework, permitting graphical specification of rules and merger of networks generated automatically (using the BioNetGen modeling engine) with hand-specified reaction networks. VCell provides a number of ordinary differential equation and stochastic numerical solvers for single-compartment simulations of the kinetic systems derived from these networks, and agent-based network-free simulation of the rules. In this work, compartmental and spatial modeling of rule-based models has been implemented within VCell. To enable rule-based deterministic and stochastic spatial simulations and network-free agent-based compartmental simulations, the BioNetGen and NFSim engines were each modified to support compartments. In the new rule-based formalism, every reactant and product pattern and every reaction rule are assigned locations. We also introduce the rule-based concept of molecular anchors. This assures that any species that has a molecule anchored to a predefined compartment will remain in this compartment. Importantly, in addition to formulation of compartmental models, this now permits VCell users to seamlessly connect reaction networks derived from rules to explicit geometries to automatically generate a system of reaction-diffusion equations. These may then be simulated using either the VCell partial differential equations deterministic solvers or the Smoldyn stochastic simulator.

INTRODUCTION

The specification of all molecular species and interactions is usually the first step in modeling a biomolecular interaction network. However, for the interactions of multivalent or multistate molecules, the number of species and reactions can be combinatorially large (1,2), making it impractical to specify the reaction network manually. Rule-based modeling (2,3) overcomes this limitation by accounting for the complete set of reactions and species that arise when an initial (seed) set of species is transformed using reaction rules. The reaction rules can serve either as generators of individual reactions, expanding the initial set of species into the complete network of reactions and species, or as generators of stochastic events, producing molecular complexes with nonzero population numbers.

Virtual Cell (VCell; http://vcell.org) is an open-source platform that provides powerful capabilities for kinetic modeling of cellular systems (4,5) (Fig. 1). A key focus of VCell is to allow modelers to ask how spatial features of cells affect the system behavior. At the simplest level, the relative sizes of compartments affect the concentrations of species transported between them; models that account for the surface areas of membranes and the volumes of volumetric compartments, but assume that diffusion is fast on the timescale of reaction kinetics, will be referred to as “compartmental”. If diffusion and spatial localization of molecular species can affect the biology, the geometric shapes of the membrane and volumetric compartments also need to be explicitly considered, and these models are considered “spatial”.

In building a VCell “BioModel”, users initially describe the system “Physiology” with compartments defined as multiple volumes (e.g., extracellular space, cytosol, nucleus, endoplasmic reticulum, etc.) and surfaces (e.g., plasma membrane, mitochondrial membrane, etc.); the Physiology also encompasses reactions and fluxes taking place within and between volumes and surfaces, with respective volume- and area-based units for concentrations and kinetic rate expressions. Once a Physiology is defined in VCell, any number of “Applications” can be defined that specify the initial conditions, compartment sizes, and/or geometries, and whether the system should be treated deterministically.
or stochastically. An Application can be considered to be a virtual experiment and is sufficient to completely define the system’s mathematical equations, which are automatically generated. VCell applications that do not include explicitly defined geometry (i.e., making the assumption that diffusion is fast on the timescale of all reaction kinetics) are called “compartmental”. These can be simulated using a variety of deterministic and stochastic numerical solvers (Fig. 1) to produce time courses for concentrations and/or population numbers of species. Alternatively, an explicit geometry can be defined using a variety of methods, such as analytic equations in 1D, 2D, or 3D, constructed solid geometry in 3D, image-based (imported from various microscopy formats), mesh-based (imported from STL image), or drawn using provided drawing tools. Such VCell applications are called “spatial”, and diffusion for species must be defined to simulate time courses. Diffusion can be defined in both volumetric compartments and along the surface of membranes, and reactions spanning multiple compartments account for species flux between these compartments. VCell offers several solvers for partial differential equations (PDEs) (4–7) to simulate spatial and temporal changes in concentration when the species copy numbers are large. When the copy numbers are low, a spatial stochastic simulator using the Smoldyn simulation engine (8) is available. VCell even has a solver for spatial hybrid deterministic/stochastic simulations (9), to accommodate systems containing some species at high copy number (modeled as continuous concentrations and PDEs) and others at low copy number (modeled with stochastic reaction kinetics and Brownian motion).

Earlier versions of VCell required explicit specification of species and reactions. Last year we introduced VCell 6.0, which incorporated a graphical user interface (GUI) to represent multiple sites and states within molecules and the rule-based reaction kinetics between them (10). This GUI provides a compact method for describing the key structural features of multivalent multistate molecules that control their roles in complex signaling systems. Every chemical species can be represented as structured objects...
composed of molecules, with reactions that control all their modifications and changes in their connectivity. Every model can be simulated both as a reaction network (after network-generation using the BioNetGen engine), permitting both deterministic and stochastic simulations, or with the NFSim network-free algorithm, which produces stochastic simulations. However, the abstractions within the representations of molecules and rules, as well as the algorithms within the network generation and simulation engines, did not include compartments or the ability to simulate reaction diffusion equations in explicit geometries.

In this article, we describe a compartmental extension of the rule-based modeling capabilities of VCell, available in VCell 6.1. It enables specification of the locations of molecules and rules. This required us to develop new abstractions to anchor molecules explicitly to volume or surface compartments. We then modified the BioNetGen code to support network generation within the VCell compartmental formalism. This permitted us to support all the stochastic or deterministic, and nonspatial or spatial, simulators available in VCell to simulate reaction networks generated by rules (Fig. 1). Additionally, we modified the NFSim code (11) to support compartmental (albeit nonspatial) network-free simulations.

MATERIALS AND METHODS

Rule-based modeling in VCell is implemented by adapting the standalone software tools BioNetGen (3,12) and NFSim (11). They allow both deterministic and stochastic simulations after the reaction network is generated (BioNetGen) and network-free particle-based simulations (NFSim) in a single compartment. Both tools operate using the BioNetGen Language (BNGL), (12), which was originally designed to work only for a single compartment. A VCell-specific BNGL extension was developed to enable model specification in multiple compartments. The BioNetGen and NFSim tools were each modified to support multiple compartments and new VCell-specific features.

RESULTS

A previously described compartmental extension of BNGL (cBNGL), (13), enables explicit modeling of the compartment topology of the cell and its effects on system dynamics using the BioNetGen network generation engine. However, it is not suited for VCell. The major reason is that cellular topology in cBNGL is restricted to a compartment graph in which nodes represent compartments and directed edges represent containment. This graph must be a tree. A membrane may contain (and be contained in) only a single volume compartment, whereas a volume compartment may contain multiple surface compartments but be contained only by a single membrane. This cellular topology is more limited than the generalized topology available in VCell, where no restrictions are imposed on how compartments can be enclosed within each other. Likewise, cBNGL does not allow for molecular species to span several membranes. The VCell paradigm gives the modeler more flexibility and supports representation of multicellular structures with gap junctions or tight junctions. Additionally, NFSim does not support compartmental simulations using cBNGL. Therefore, to enable rule-based modeling in a generalized topology using both the BioNetGen and NFSim engines, we decided to develop our own schema based on the VCell multicompartment reaction diagram, where the locations of each reactant, product, and of the reaction itself are explicitly specified. This required new conventions for specifying spatial features within a regular BNGL file. We use the BioNetGen or NFSim engines to generate the reactions, but have added a processing step after every iteration to fix the location of each newly generated species and remove invalid species and reactions. Although the BNGL is hidden to the user, it is important to lay out the key algorithmic features that allow for merging of rule-based modeling within the VCell architecture. We describe these in Fig. 2 and Table 1, which also serve as a short glossary of rule-based modeling terminology. Note that although a species is located in a given compartment, its orientation in space (as may be required for agent-based simulations) is implicitly determined by the specification of binding reaction rules between sites on molecules with different locations. Supporting Material discusses implementation in more detail.

Let us illustrate compartmental rule-based modeling using a model of Ran-mediated nucleocytoplasmic transport (available in the VCell Database under Tutorials as “Rule-based_Ran_transport”). This is a simplified version of a published model by Smith et al. (14). In this tutorial (Fig. 3 A), the nuclear Ran binds to a cargo molecule, facilitating translocation into the cytosol. Ran and cargo then dissociate. The cytosolic cargo molecule may be phosphorylated on any of its three tyrosines while in the cytosol. When cargo is displaced by the Ran exchange factor RCC1, which is a component of histones, Ran stays in the nucleus. The rule-based approach provides a compact way of describing such a system, with a single transport reaction rule (Fig. 2 E and rule 7 in Fig. 3 A) in place of 18 reversible reactions for transport of multiple types of cargo. The total number of species and reactions, if all combinations of phosphoforms are accounted for, would be 36 and 98, respectively. Such a large system is difficult to construct, visualize, and analyze without rule-based specification (Fig. 3 B). Moreover, model predictions have to be compared to experimental observations that often correspond to sums of multiple species. In our example, the total phosphorylation of cargo is summed over all species that have at least one site being phosphorylated, with triple-phosphorylated species counted three times. The observable concept (Fig. 1 D and Table 1) introduced in rule-based modeling provides an easy way to define such quantities.
A feature of rule-based modeling is that anything not explicitly constrained is allowed. Thus, a transport reaction rule for Ran with a pattern specifying it must have a bound site (Fig. 2 E), will carry with it any of the molecules that may be bound to it, including RCC1, which should remain in the nucleus according to the known biology. To address this within the generalized topology of VCell, we introduced the ability to “anchor” molecules to compartments (Fig. 4). A molecule can be anchored to one or multiple compartments. Anchoring a molecule to specific compartments constrains any species (produced through network generation or by NFSim) that contains this molecule to be in only these compartments. In spatial applications, such species are free to diffuse within compartments they are anchored to.

As summarized in the Introduction, VCell has a hierarchical architecture whereby the system “Physiology” (exemplified by the network in Fig. 3 and all its underlying details) can be associated with several Applications. Applications are used to specify initial conditions, geometries, and the physical and mathematical approaches by which the system should be simulated. New Application types have been developed to provide control for network generation and to support a network-free simulation with NFSim. All the simulation methods previously available to manually constructed reaction networks are thereby now available to networks generated automatically from rules. Fig. 5 shows how the Physiology summarized in Fig. 3 can be used to produce compartmental deterministic (ordinary differential equations), compartmental stochastic, compartmental network-free, spatial deterministic (PDE), spatial stochastic, and spatial hybrid deterministic/stochastic simulations. We believe that VCell is unique in making all these approaches available within one unified software.
environment; we are also unaware of any other systems biology software offering a spatial hybrid deterministic/stochastic solver (9). But VCell does not have pretensions of being a solution for every modeling problem. Accordingly, we have devoted significant effort to interoperability—in particular through the SBML standard (15). Specifically, for the case of rule-based modeling, a model defined in VCell can be exported to cBNGL (13) for simulation with the stand-alone BioNetGen engine. The generated cBNGL file has a new “anchors block” specifying molecules anchored to compartments. This block is ignored by the BioNetGen compiler. The exported cBNGL has no information about enclosing compartments in the compartment block. Thus,

<table>
<thead>
<tr>
<th>Term</th>
<th>Purpose</th>
<th>Composition</th>
<th>Compartmental/Spatial Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compartment</td>
<td>cellular structure containing species and reactions</td>
<td>a volume or a surface, corresponding to extracellular regions, cytosol, and organelles and their associated membranes; no specification of the relationship between compartments (such as enclosures or adjacency) is required</td>
<td>for Applications without explicit geometries, compartmental sizes are specified as surface areas or volumes; for spatial Applications, compartments may be explicitly associated with regions within a geometry</td>
</tr>
<tr>
<td>Molecule (Fig. 2 A)</td>
<td>a building block for species</td>
<td>composed of sites that can bind other sites between or within molecules; a site may also have multiple states; in this way, a molecule spawns a collection of chemical species—one per every combination of site occupancy and/or state</td>
<td>can be, optionally, “anchored” to one or more compartments; a species containing an anchored molecule can be located only in one of these compartments</td>
</tr>
<tr>
<td>Species (Fig. 2 B)</td>
<td>an individual chemical species that may occur in the model</td>
<td>composed of molecules that are connected through bonds between binding sites; all modification sites must be explicitly defined; a species can be a “seed species” (defined as initial condition and used as a seed for reaction rules application) or a generated species (a result of reaction rule application)</td>
<td>every species is located in a unique compartment; seed species are assigned to a compartment by the user; it may not contain molecules that are excluded from that compartment by anchoring to other compartments</td>
</tr>
<tr>
<td>Pattern (Fig. 2 C)</td>
<td>specifies a set of possible states of species to be selected as participants in reaction rules and in observables</td>
<td>composed of molecules; the states of sites may be left unspecified; thus a pattern may select multiple species; moreover, binding sites may have “implicit binding status” (has external bond or may be bound) where its binding partner is not explicitly defined; such patterns may be inclusive of species that contain molecules not explicitly specified in a pattern but being possibly bound to molecules within it</td>
<td>defined in a single compartment; all molecules comprising a pattern must be permitted to be located in this compartment</td>
</tr>
<tr>
<td>Observable (Fig. 2 D)</td>
<td>specifies simulation outputs of interest</td>
<td>consists of one or more patterns that define features of species; the result is the total population (concentration or count) of multiple species</td>
<td>defined in a single compartment; all molecules comprising an observable must be permitted to be located in this compartment</td>
</tr>
<tr>
<td>Reaction rule (Fig. 2 E)</td>
<td>defines transformation of multiple species at once, generating multiple reactions</td>
<td>species to be transformed are selected by reactant pattern(s); product pattern(s) define the end result of transformation; product may differ from reactant by reassigning molecules, adding or deleting bonds and changing site states; a kinetic expression is also a component of a reaction rule</td>
<td>reaction rule is defined in a single compartment; each reactant pattern and each product pattern are assigned a specific compartment, which may be different from compartment for reactants or products</td>
</tr>
</tbody>
</table>
to simulate an exported model with the standalone BioNetGen engine, a modeler needs to specify a compartmental tree manually. Also, a model specified in cBNGL can be imported into VCell. Not every file can be seamlessly imported; errors will be displayed when compartment specification is done at a level of individual molecules in seed species and patterns. VCell provides a BNGL import editor where all inconsistencies are displayed and explained, so an experienced BioNetGen user should be able to fix all issues during the import process. Supporting Material provides more details on comparison between cBNGL and VCell representation.

DISCUSSION

We have described a major enhancement of the VCell software to enable rule-based modeling in multiple compartments. This enhancement gives users with combinatorially complex biochemical systems the ability to specify all interactions and their dependencies in terms of molecular features such as cellular locations, sites for binding, modification states, or conformations. To achieve this, we built a rich GUI that also serves to help visualize the details of these complex systems (Fig. 3). The VCell “classic” manual network generation functionality and GUI are still available and the implementation actually supports mixing of automatically generated rule-based networks with reaction networks generated manually. Such networks can then be modeled with all the VCell compartmental and spatial simulation methods (Fig. 5). For network-free simulations, we have also modified the NFSim engine to support compartments.

There are limitations to rule-based modeling that users should appreciate. One is the restriction of reaction kinetics to mass action; however, the VCell user may be able to overcome this restriction for deterministic models by judiciously mixing rules and explicit network reactions. A second restriction is that the network-free simulations can only be run for nonspatial models. Additionally, for spatial models based on reaction networks generated by rules, care needs to be exercised not to allow a combinatorial explosion of species and reactions; the resultant large system of PDEs could produce prohibitively expensive computations.

Several rule-based modeling tools that can operate in multiple compartments or perform spatial simulations are
available: Simmune (16,17), KaSim (18), Smoldyn (8), 
Meredys (19), SRSim (20), SSC (21), and SpringSaLaD 
(22). All of them are exclusively stochastic, whereas VCell 
offers deterministic spatial and hybrid deterministic/ 
stochastic simulation capabilities. The VCell spatial sto-
chastic solver is based on Smoldyn (23), adapted to permit 
users to incorporate experimental 3D image-based geome-
tries in simulations; analytical geometries and constructive 
geometries can also be used. A variety of nonspatial 
compartmental simulators, both stochastic and determin-
istic, are also in VCell for quick answers when diffusion 
is fast on the timescale of reaction kinetics. Additionally, 
only Simmune and SpringSaLaD have biology oriented 
GUIs, as in VCell, whereas all the other simulators are based 
on scripting. But the other more specialized simulators have 
important strengths that might be needed for certain classes 
of problems. SSC and Smoldyn both have implementations 
to employ high performance computing or graphics process-
ing units for computationally intensive simulations. Sim-
mune is specialized for complex signaling in immunology. 
Meredys, SRSim, and SpringSaLaD are all designed to 
count for molecular-excluded volume effects and are 
therefore well suited for simulations where molecular 
crowding might be important. To facilitate interoperability 
with such other simulators, VCell supports the SBML stan-
dard (15) by enabling export and import of SBML models; it 
also supports cBNGL export, although some manual editing 
will be required.

Model sharing is also facilitated within VCell through the 
VCell database. All models can be stored in the database 
along with simulations results that were run on the VCell 
server farm (although users may opt to save models and 
run simulations on their local machines). Access control is 
implemented to permit sharing of models with individual 
collaborators or to make a model openly accessible. Users 
may annotate model components to connect them to the 
primary literature sources as well as to ontologies and 
pathway databases. This is particularly valuable for mole-
cules in rule-based models, where the localization and sites 
within a molecule can be directly related to both molecular 
structure and pathway data. Importantly, proper annotation 
can assure reusability of not just the entire model, but the 
individual molecules and rules.

Software availability

This software is available as VCell (versions 6.1 and later) at 
the Virtual Cell web site (http://vcell.org/). The application
installs and runs on all major platforms and does not require registration for use on the user’s computer. Tutorials are available at the Virtual Cell website and Help is provided within the software. Source code is available at https://github.com/virtualcell.

SUPPORTING MATERIAL

Supporting Materials and Methods, fourteen figures, one movie, and one QuickStart Guide are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(17)30915-3.

ACKNOWLEDGMENTS

The authors gratefully acknowledge useful discussions with James Faeder at the University of Pittsburgh and Ann Cowan at UConn Health. The project is supported by the National Institute of General Medical Sciences (NIGMS) through grants P41 GM103313 and R01 GM095485.

REFERENCES