

Compartmental and spatial rule-based modeling with Virtual cell (VCell)

M. L. Blinov,* J. C. Schaff, D. Vasilescu, I. I. Moraru, J. E. Bloom and L. M. Loew*

Running title: Compartmental rule-based modeling in VCell

Keywords: Virtual Cell (VCell), rule-based modeling, compartments

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 (ODE) 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 novel 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 (PDE) 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 non-zero 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). 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 reactions occurring within volumetric or membrane compartments, and including transport mechanisms between compartments. Once a 'Physiology' is defined in VCell, any number of 'Applications' can be defined, which specify the initial conditions, compartment sizes and/or geometries, and whether the system should be treated deterministically or stochastically. An 'Application' can be considered virtual experiments and are sufficient to completely define the system equations, which are automatically generated. These can then be simulated using a variety of numerical solvers: compartmental or explicit spatial geometries combined with multiple choices of deterministic or stochastic solvers. VCell even has a solver for spatial hybrid deterministic/stochastic simulations (6), to accommodate systems containing some species at high copy number (modeled as continuous concentrations and partial differential equations) and others at low copy number (modeled with stochastic reaction kinetics and Brownian motion).

Previous versions of VCell (through v. 5.3) required explicit specification of species and reactions, although a plugin with BioNetGen engine could be used to generate reaction networks to import and simulate by VCell. 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 (7). 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 (following network-generation using the BioNetGen engine), permitting both deterministic or stochastic simulations, and using network-free algorithm implemented by the NFSim engine (8), 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 paper 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 non-spatial or spatial, simulators available in VCell to simulate reaction networks generated by rules. Additionally, we modified the NFSim code (8) to support compartmental (albeit non-spatial) network-free simulations.

Results

Rule-based modeling in VCell is implemented by adapting the standalone software tools BioNetGen (3, 9) and NFSim (8). 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 (9), which was originally designed to work only for a single compartment.

A compartmental extension of BNGL, cBNGL (10), 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

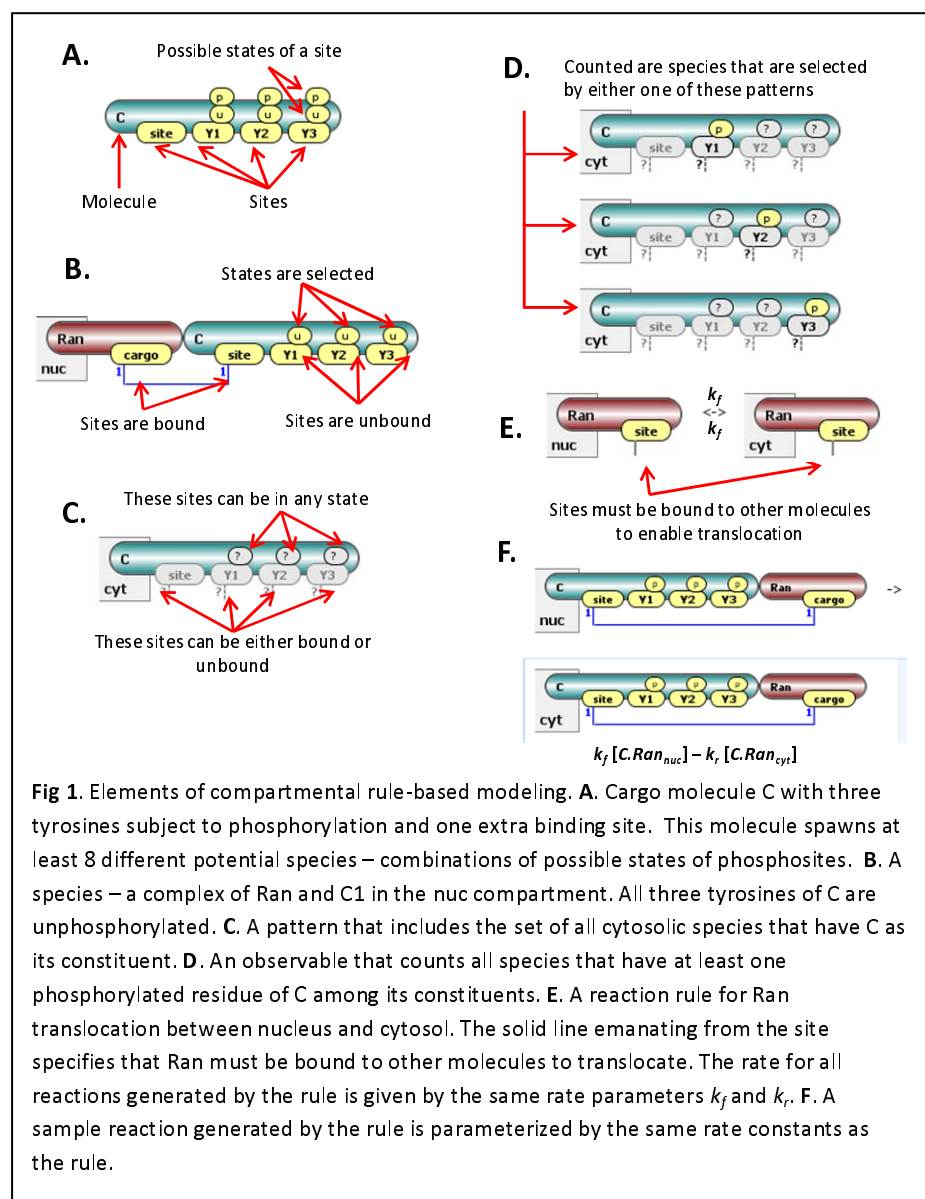


Fig 1. Elements of compartmental rule-based modeling. **A.** Cargo molecule C with three tyrosines subject to phosphorylation and one extra binding site. This molecule spawns at least 8 different potential species – combinations of possible states of phosphosites. **B.** A species – a complex of Ran and C1 in the nuc compartment. All three tyrosines of C are unphosphorylated. **C.** A pattern that includes the set of all cytosolic species that have C as its constituent. **D.** An observable that counts all species that have at least one phosphorylated residue of C among its constituents. **E.** A reaction rule for Ran translocation between nucleus and cytosol. The solid line emanating from the site specifies that Ran must be bound to other molecules to translocate. The rate for all reactions generated by the rule is given by the same rate parameters k_f and k_r . **F.** A sample reaction generated by the rule is parameterized by the same rate constants as the rule.

Table 1. Concepts and definitions for compartmental rule-based modeling in VCell

| Term | Purpose | Composition | Compartmental/Spatial features |
|-------------------------|--|--|---|
| Compartment | Cellular structure containing species and reactions | A volume or a surface, corresponding to extracellular regions, cytosol, organelles and their associated membranes. No specification of the relationship between compartments (such as enclosures or adjacency) is required. | 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. |
| Molecule (Fig 1A). | A building block for species. | Comprised 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. | Can be, optionally, anchored to one or more compartments. A species containing an anchored molecule can be located only in one of these compartments. |
| Species (Fig 1B). | An individual chemical species that may occur in the model. | Comprised 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). | 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. |
| Pattern (Fig 1C). | Specifies a set of possible states of species to be selected as participants in reaction rules and in observables. | Comprised 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 (<i>has external bond or may be bound</i>) 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. | Defined in a single compartment; all molecules that comprise a pattern must be permitted to be located in this compartment. |
| Observable (Fig 1D). | Specify simulation outputs of interest. | Consists of one or more patterns that define features of species. The result is the total population (concentration or count) of multiple species. | Defined in a single compartment; all molecules that comprise an observable must be permitted to be located in this compartment. |
| Reaction rule (Fig 1E). | Defines transformation of multiple species at once, generating multiple reactions | 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 re-assigning molecules, adding or deleting bonds and changing site states. A kinetic expression is also a component of a reaction rule. | 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. |

major reason is that cellular topology in cBNGL is restricted to a compartment graph in which nodes represent compartments and directed edges represent containership. This graph must be a tree, a

membrane may contain (and be contained in) only a single volume compartment, while 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 multi-compartment 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 r to fix the location of each newly generated species and remove invalid species and reactions. While 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 Table 1 and Figure 1, which also serve as a short glossary of rule-based modeling terminology. Note that while 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. Supplemental material (S1.pdf) discusses implementation in more detail.

Let us illustrate compartmental rule-based modeling using a model of Ran nucleocytoplasmic transport (available in the VCell Database under Tutorials as “Rule-Based_Ran_Transport”). This is a simplified adapted version of a published model by Smith et al. (11). Ran can move several types of cargo (Fig 2) between

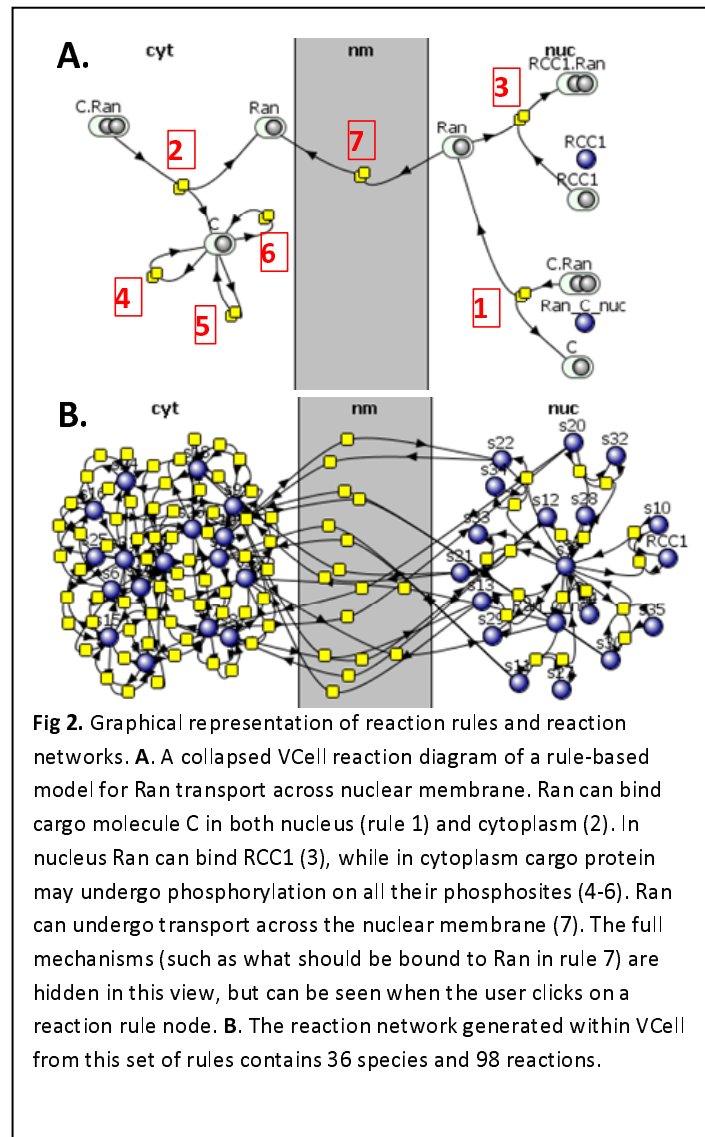


Fig 2. Graphical representation of reaction rules and reaction networks. **A.** A collapsed VCell reaction diagram of a rule-based model for Ran transport across nuclear membrane. Ran can bind cargo molecule C in both nucleus (rule 1) and cytoplasm (2). In nucleus Ran can bind RCC1 (3), while in cytoplasm cargo protein may undergo phosphorylation on all their phosphosites (4-6). Ran can undergo transport across the nuclear membrane (7). The full mechanisms (such as what should be bound to Ran in rule 7) are hidden in this view, but can be seen when the user clicks on a reaction rule node. **B.** The reaction network generated within VCell from this set of rules contains 36 species and 98 reactions.

nucleus and cytoplasm. However, 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 1E) in place of 18 reversible reactions for transport of multiple types of cargo. However, a transport reaction rule for Ran with a pattern specifying it must have a bound site (Fig 1E), will carry with it any of the molecules that may be bound to it, including RCC1. To address this within the generalized topology of VCell, we introduced the ability to *anchor* molecules to compartments (Fig 3). A molecule can be anchored to one or more compartments.

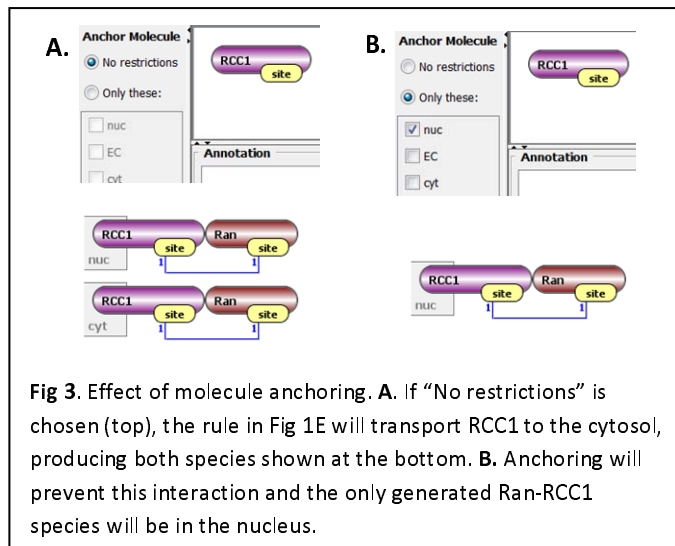


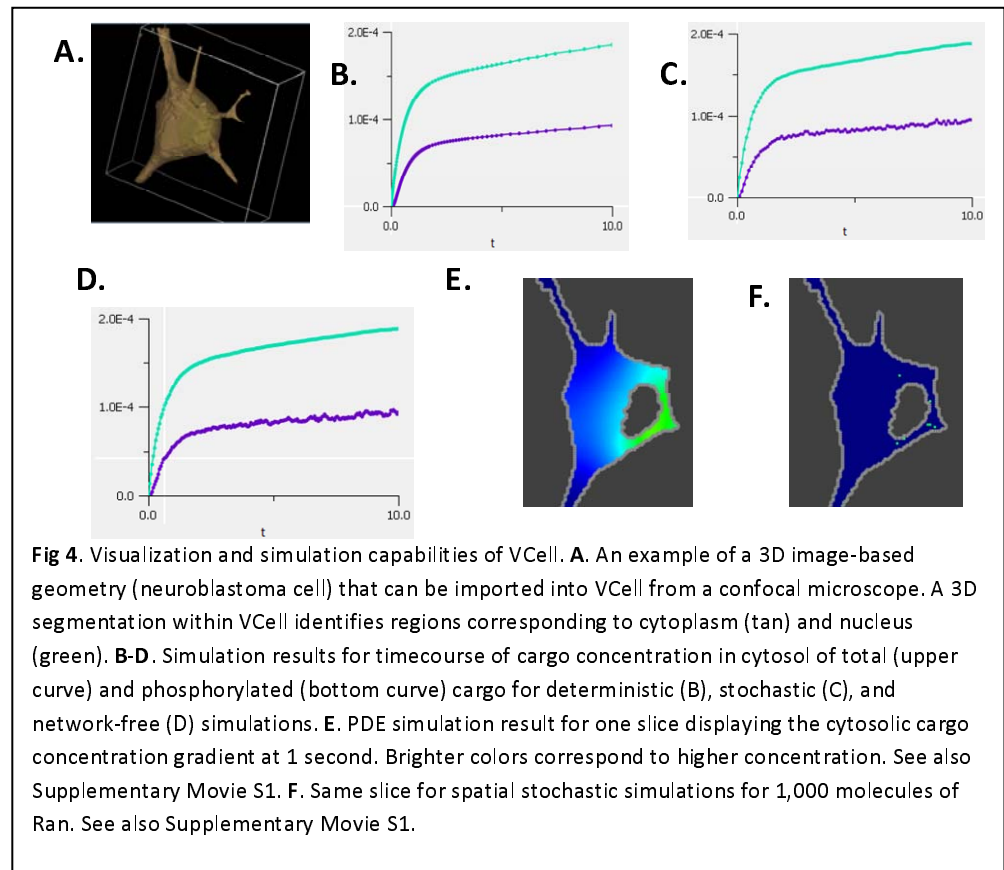
Fig 3. Effect of molecule anchoring. **A.** If “No restrictions” is chosen (top), the rule in Fig 1E will transport RCC1 to the cytosol, producing both species shown at the bottom. **B.** Anchoring will prevent this interaction and the only generated Ran-RCC1 species will be in the nucleus.

Anchoring a molecule to specific compartments restricts any species (produced through network generation or by NFSim) that contains this molecule to be in only these compartments.

As summarized in the introduction, VCell has a hierarchical architecture whereby the system ‘Physiology’ (exemplified by the network in Fig. 2 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.

Applications have been enhanced to provide control for network generation and to support a special Application type for network-free simulations with NFSim. Now that the rule-based VCell Physiology has been extended to support compartments, all the simulation methods previously available to manually constructed reaction networks are available to networks generated automatically from rules. Fig. 4 shows how the Physiology summarized in Fig. 2 can be used to produce compartmental deterministic (ODE), 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(6).

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 (12). Specifically for the case of rule-based modeling, a model defined in VCell can be exported to cBNGL (10) 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, to simulate an exported model with the stand-alone 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 in VCell editor during the import process. Supplemental material S2.pdf 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. 2). 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. 4). For network-free simulations, we have also modified the NFSim engine to support compartments.

Several rule-based modeling tools that can operate in multiple compartments or perform spatial simulations are available: Simmune (13, 14), KaSim (15), Smoldyn (16), Meredys (17), SRSim (18), SSC (19), SpringSaLaD (20). All of them are exclusively stochastic, whereas VCell offers deterministic spatial and hybrid deterministic/stochastic simulation capabilities. The VCell spatial stochastic solver is based on Smoldyn (21), adapted to permit users to incorporate experimental 3D image-based geometries in simulations; analytical geometries and constructive geometries can also be used. A variety of non-spatial compartmental simulators, both stochastic and deterministic, 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, while 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 gpus for computationally intensive simulations. Simmune is specialized for complex signaling in immunology. Meredys, SRSim and SpringSaLaD are all designed to account 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 standard (12) 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 molecules 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

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://sourceforge.net/projects/vcell/>

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

Supporting Material

Supplemental text 1 describes materials and methods. Supplemental text 2 contains a quick software guide. Supplemental movies describe simulation results for spatial deterministic and stochastic applications.

References

1. Mayer, B. J., M. L. Blinov, and L. M. Loew. 2009. Molecular machines or pleiomorphic ensembles: signaling complexes revisited. *J Biol* 8:81.81-81.88.
2. Faeder, J. R., M. L. Blinov, B. Goldstein, and W. S. Hlavacek. 2005. Combinatorial complexity and dynamical restriction of network flows in signal transduction. *Syst Biol (Stevenage)* 2:5-15.
3. Blinov, M. L., J. R. Faeder, B. Goldstein, and W. S. Hlavacek. 2004. BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* 20.
4. Cowan, A. E., Moraru, II, J. C. Schaff, B. M. Slepchenko, and L. M. Loew. 2012. Spatial modeling of cell signaling networks. *Methods Cell Biol* 110:195-221.
5. Schaff, J. C., C. C. Fink, B. Slepchenko, J. H. Carson, and L. M. Loew. 1997. A general computational framework for modeling cellular structure and function. *Biophysical journal* 73:1135-1146.
6. Schaff, J. C., F. Gao, Y. Li, I. L. Novak, and B. M. Slepchenko. 2016. Numerical Approach to Spatial Deterministic-Stochastic Models Arising in Cell Biology. *PLoS Comput Biol* 12:e1005236.
7. Schaff, J. C., D. Vasilescu, I. I. Moraru, L. M. Loew, and M. L. Blinov. 2016. Rule-based Modeling with Virtual Cell. *Bioinformatics* 32:2880-2882.
8. Sneddon, M. W., J. R. Faeder, and T. Emonet. 2011. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Methods* 8:177-183.
9. Faeder, J. R., M. L. Blinov, and W. S. Hlavacek. 2009. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol Biol* 500:113-167.
10. Harris, L. A., J. S. Hogg, and J. R. Faeder. 2009. Compartmental rule-based modeling of biochemical systems. In *Winter Simulation Conference*. 908-919.
11. Smith, A. E., B. M. Slepchenko, J. C. Schaff, L. M. Loew, and I. G. Macara. 2002. Systems analysis of Ran transport. *Science* 295:488-491.
12. Hucka, M., A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, Goryanin, II, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama,

- M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524-531.
13. Angermann, B. R., F. Klauschen, A. D. Garcia, T. Prustel, F. Zhang, R. N. Germain, and M. Meier-Schellersheim. 2012. Computational modeling of cellular signaling processes embedded into dynamic spatial contexts. *Nat Methods* 9:283-289.
 14. Zhang, F., B. R. Angermann, and M. Meier-Schellersheim. 2013. The Simmune Modeler visual interface for creating signaling networks based on bi-molecular interactions. *Bioinformatics* 29:1229-1230.
 15. Sorokina, O., A. Sorokin, J. Douglas Armstrong, and V. Danos. 2013. A simulator for spatially extended kappa models. *Bioinformatics* 29:3105-3106.
 16. Andrews, S. S. 2017. Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. *Bioinformatics* 33:710-717.
 17. Tolle, D. P., and N. L. Novère. 2010. Meredys, a multi-compartment reaction-diffusion simulator using multistate realistic molecular complexes. *BMC Syst Biol.* 4.
 18. Gruenert, G., B. Ibrahim, T. Lenser, M. Lohel, T. Hinze, and P. Dittrich. 2010. Rule-based spatial modeling with diffusing, geometrically constrained molecules. *BMC Bioinformatics* 11.
 19. Lis, M., M. N. Artyomov, S. Devadas, and A. K. Chakraborty. 2009. Efficient stochastic simulation of reaction-diffusion processes via direct compilation. *Bioinformatics* 25:2289-2291.
 20. Michalski, P. J., and L. M. Loew. 2016. SpringSaLaD: A Spatial, Particle-Based Biochemical Simulation Platform with Excluded Volume. *Biophysical journal* 110:523-529.
 21. Andrews, S. S., N. J. Addy, R. Brent, and A. P. Arkin. 2010. Detailed simulations of cell biology with Smoldyn 2.1. *PLoS computational biology* 6:e1000705.