

Modeling Complex Networks: Integrating Rules (BioNetGen) and Data Mining (BioPAX Ontology) into the Virtual Cell Framework

Ion I. Moraru, Oliver Ruebenacker, James C. Schaff,
Leslie M. Loew and Michael L. Blinov
Center of Cell Analysis and Modeling,
University of Connecticut Health Center, Farmington, CT
E-mail: blinov@uchc.edu

A combinatorial complexity often arises when detailed quantitative models of intracellular networks are being sought. Proteins are composed of functional modules (e.g., Src homology 2 (SH2) binding sites and tyrosines phosphorylation sites [1]). A receptor that has 9 tyrosine phosphorylation sites can exist in $2^9 = 512$ different phosphoforms. Many or all of these phosphoforms must be accounted for in a quantitative model to simulate the time course for receptor-mediated signaling [2]. Simple models of biochemical kinetics accounting for dozens of different molecular species are a norm; models accounting for hundreds of species and reactions are no longer rare. When details of all functional forms are being included, this number can easily increase by a few orders of magnitude, and validation, visualization, and understanding can become virtually intractable (Fig. 1a). A solution for this challenge is provided by 1) automatic extraction from pathway databases re-usable model components for quantitative models, and 2) rules of interaction based on protein modularity (e.g. kinetics of binding is independent of other sites). This way, quantitative models of large, complex networks can be assembled from separately constructed, validated, and visualized components, either directly or via rules.

To implement this strategy, we have combined the strength of several related technologies: the Biological Pathways Exchange (BioPAX) ontology (<http://biopax.org>), the Systems Biology Markup Language (SBML) format (<http://sbml.org>), the BioNetGen rule-based description of molecular interaction (<http://bionetgen.lanl.gov>), and the Virtual Cell (VCell) [3, 4] modeling and simulation software framework (<http://vcell.org>). Two approaches are used to generate models without manual specification of each and every species and reactions. First is using BioPAX data imported from BioPAX-compatible databases, e.g. Reactome [5]. Each element of a BioPAX file is linked to an originating biological database, providing for a well-documented biological identification for each element of the model. Any species and reaction annotated with BioPAX can be easily encapsulated in a reusable modeling module. The application automatically generates an SBML file that can be simulated after kinetic parameters are added by the modeler. This also allows for better visualization of the model, with different types of species and reactions having different notations (Fig 1b). Second is to specify a model in the form of bio-molecular interaction rules that generate a biochemical reaction network [6]. This approach has been implemented in general-purpose software, BioNetGen [7], and recently has been implemented as a BioNetGen@VCell application. In this approach, a modeler specifies molecules as functional modules consisting of multiple elements (interacting and modification sites), and rules of activities and interactions among these elements (Fig 2), which are automatically converted into a reaction network that can be shared with other tools via SBML

[8]. However, this does not provide exchange of the actual models, as essential features that were used to generate the reaction networks (e.g. molecules, their components, and rules) are lost. Thus, an extension of SBML is required to allow an exchange of models accounting for modular structure of models and their components.

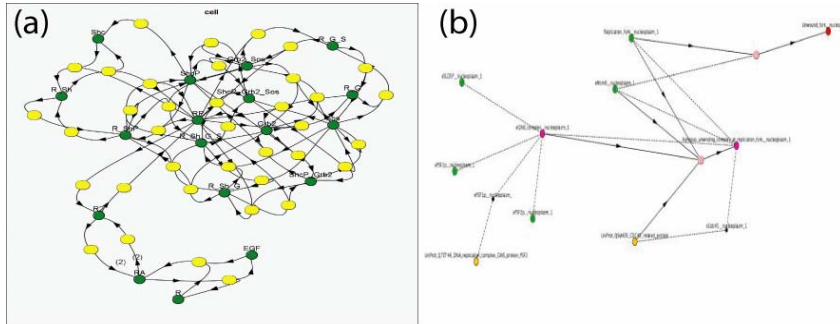


Figure 1 (a) The screenshot of VCell reaction network. The explicit representation of each and every species and reaction is difficult to visualize. (b) The screenshot of BioPAX@VCell representation, with different coloring and symbols for different types of species and reactions.

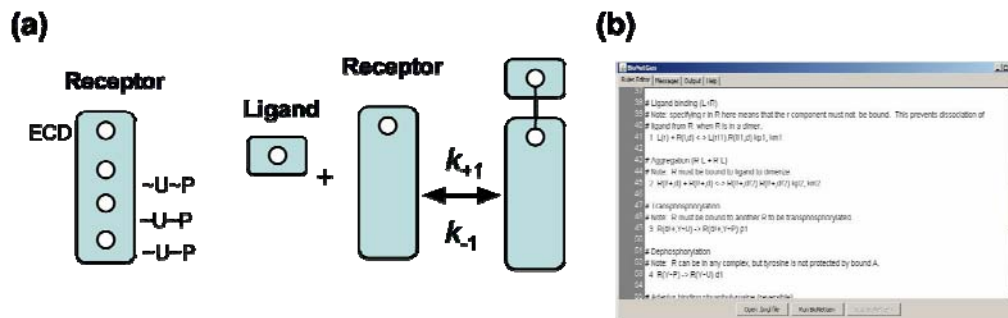


Figure 2 (a) Rules are based on the knowledge of modular structure of proteins, such as interaction of ligand with extracellular domain of the receptor is independent from the state of intracellular tyrosines. Here, a receptor consists of 4 elements (domains): extracellular ECD and 3 phosphosites that can be in two states (U and P). Thus, the total number of potential phosphoforms of this receptor is $2^3=8$. However, the rule does not specify a state of any of phosphosites, which implies the same rate of ligand-receptor binding for each of 8 potential phosphoforms. (b) The screenshot of rules representation in the BNG@VCell application. 5 rules generate 110 individual reactions among 16 chemical species.

References:

- [1] Pawson T, Nash P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*. 300:445-52.
- [2] Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. (2006) A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems*. 83:136-51.
- [3] Loew LM, Schaff JC. (2001) The Virtual Cell: A software environment for computational cell biology. *Trends in Biotechnology*, 19:401–406.

- [4] Moraru II, Schaff JC, Loew LM (2006). Think simulation – think experiment: the Virtual Cell paradigm. In Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto RM, eds., Proceedings of the 2006 Winter Simulation Conference, pp. 1713-1719.
- [5] Vastrik I, D'Eustachio P, Schmidt E, Joshi-Tope G, Gopinath G, Croft D, de Bono B, Gillespie M, Jassal B, Lewis S, Matthews L, Wu G, Birney E, Stein L. (2007) Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.* 8(3):R39.
- [6] Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M, Fontana W. (2006) Rules for modeling signal-transduction systems. *Sci STKE.* 2006(344):re6.
- [7] Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. (2004) BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics.* 20:3289-91
- [8] Hucka M et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics.* 19:524-31.