Published in final edited form as:

Nat Methods. 2011 February ; 8(2): 130–131. doi:10.1038/nmeth0211-130.

New approaches to modeling complex biochemistry

John A Bachman and Peter Sorger

John A. Bachman and Peter Sorger are in the Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA

Peter Sorger: peter_sorger@hms.harvard.edu

Abstract

Combining rule-based descriptions of biochemical reactions with agent-based computer simulation opens new avenues for exploring complex cellular processes.

A major challenge in modeling complex genetic and biochemical circuits is turning imprecise ‘word models’ (text and drawings) into precise mathematical statements suitable for computational analysis. Many biochemical processes common to eukaryotic signaling networks are remarkably difficult to describe and simulate with rigor and precision. Assembly of multiprotein complexes or multisite post-translational modification can create a very large number of distinct molecular species from interactions involving only a handful of gene products. Such ‘combinatorial complexity’ exceeds the ability of conventional modeling methods to enumerate species and track their dynamics, leading to *ad hoc* assumptions about which types of interactions matter and which do not.

The network-free stochastic simulator (NFSim), described in this issue of *Nature Methods*¹, is a new entry in an evolving class of general-purpose computational tools that addresses the challenge of combinatorial complexity by combining a rule-based representation with agent-based simulation. In a rule-based representation, patterns of interaction among proteins and other biomolecules are specified using a specialized computer language. By analogy to organic chemistry, only information important for a specific reaction is included in a rule; all nonparticipating structures are left unspecified. A rule-based approach differs from conventional reaction-centric models in which equations (commonly ordinary differential equations (ODEs)), one for every possible molecular species or complex, are enumerated to describe the time dynamics of the system. For combinatorially complex systems, equation-based models are hard to error-check, extend and reuse, in contrast to rule-based models, which are concise, comprehensible and easily extended. Research to date suggests that rule-based approaches enable simulation and analysis of classes of complex reactions that would otherwise be intractable².

Available rule-based modeling tools are differentiated by the way they perform simulations (Fig. 1). In BioNetGen³, the technology on which NFSim is built, a rule set is typically used to generate the full network of all possible chemical reactions, which are then simulated using ODEs or the stochastic simulation algorithm of Doob-Gillespie⁴. However, for systems with a sufficiently high degree of combinatorial complexity, enumeration of all chemical reactions becomes prohibitively costly (as the authors¹ of NFSim demonstrate, this can occur in simulations of multisite phosphorylation with as few as six to eight sites).

To overcome this limitation, NFSim represents the system as a finite pool of interacting molecules, or ‘agents’, and the simulation unfolds stochastically by the repeated action of the rules on the pool of molecules. Because the rules themselves are used to direct the simulation, generation of the full reaction network is unnecessary. In this respect, NFSim adopts an approach pioneered by Kappa, an alternative but closely related rule-based modeling tool that also uses a network-free simulation algorithm⁵. The NFSim authors¹ demonstrate increased simulation efficiency over Kappa and other existing rule-based tools for a small set of example models, an improvement attributed to extensive software optimization.

Although rules are more intuitive and concise than systems of differential equations, existing rules have limited expressive power. For example, if the activity of a multisubunit enzyme varies with conformation, a set of similar rules is needed for each distinct configuration of the enzyme. NFSim introduces ‘local functions’, a feature in which the rate of a reaction rule can depend on the properties of a protein complex in arbitrarily definable ways reflective of empirical data. The authors’ application of NFSim to the study of bacterial chemotaxis is a useful example of the benefits of this approach¹ as their model could not be described using either a strictly ODE- or rule-based approach.

Quantitative information on even the best-understood biochemical pathways is poor, and practical models of real biology usually include both mechanistic and nonmechanistic aspects. The functional rate laws in NFSim allow users to encode the kinetic relationships between model species in terms of arbitrary mathematical functions, including Boolean functions and conditional expressions (for example, ‘if inhibitor is off, turn synthesis of your favorite gene on’) that often correspond to current levels of experimental understanding.

The ability of NFSim to blend a mechanistic representation of molecular interactions with mechanistically agnostic expressions enables ‘modeler-driven coarse-graining’. For example, a precise model of cyclin degradation through the cell cycle could be embedded in a rule-based model without including detailed information on the biochemistry of protein ubiquitination; the actual degradation machinery would be simply represented as an abstract degradation equation. This approach differs from what could alternatively be called ‘model-driven coarse-graining’, also known as model reduction, in which formal, analytical or numerical techniques are applied to a detailed model to derive a set of simplifying assumptions that allow the dynamics of the original model to be retained.

The efficiency and multiresolution modeling features offered by NFSim, in combination with its accompanying Matlab tools for parameter estimation and analysis of model output, make it a valuable general-purpose addition to the tool-box of the biological modeler. However, it is not without limitations: whereas the agent-based simulation approach adopted by NFSim tames the problem of combinatorial complexity by obviating the need to generate the entire reaction network, this approach, as all agent-based approaches, is computationally very costly when the number of molecules, or more precisely the number of molecular interactions, is high (greater than $\sim 10^4$ – 10^6 molecules of each species, using current technology). Also, the use of NFSim’s features for multiresolution modeling may preclude the application of powerful model checking⁶ and model reduction algorithms that have recently been described for ODE-based⁷ and rule-based^{8,9} models. Additional research is also required to determine the extent to which the flexible, hybrid models that can be encoded in NFSim remain analytically tractable.

Rapid progress in the development of NFSim, Kappa, little b¹⁰ and related metalanguages for representing biochemical reactions are the harbingers of an entirely new way of

representing and studying cellular networks. We predict that within a decade these methods will be mainstream components of modern quantitative biology.

References

1. Sneddon MW, Faeder JR, Emonet T. *Nat Methods*. 2011; 8:177–183. [PubMed: 21186362]
2. Hlavacek WS, et al. *Sci STKE*. 2006; 344:re6. [PubMed: 16849649]
3. Faeder JR, Blinov ML, Hlavacek WS. *Methods Mol Biol*. 2009; 500:113–167. [PubMed: 19399430]
4. Gillespie DT. *J Comput Phys*. 1976; 22:403–434.
5. Danos V, Feret J, Fontana W, Krivine J. *Lect Notes Comput Sci*. 2007; 4807:139–157.
6. Danos V, Feret J, Fontana W, Krivine J. *Lect Notes Comput Sci*. 2008; 4905:83–97.
7. Conzelmann H, Fey D, Gilles ED. *BMC Syst Biol*. 2008; 2:78. [PubMed: 18755034]
8. Borisov NM, Chistopolsky AS, Faeder JR, Kholodenko BN. *IET Syst Biol*. 2008; 2:342–351. [PubMed: 19045829]
9. Feret J, Danos V, Krivine J, Harmer R, Fontana W. *Proc Natl Acad Sci USA*. 2009; 106:6453–6458. [PubMed: 19346467]
10. Mallavarapu A, Thomson M, Ullian B, Gunawardena J. *J R Soc Interface*. 2009; 6:257–270. [PubMed: 18647734]

Figure 1. Modeling the assembly of a three-protein complex, ABC, using a rule-based approach. A and C bind B at independent sites (sites a and c, respectively), allowing the assembly process to be represented using only two rules. When the chemical reaction network is generated, these rules expand to four chemical reactions and six ODEs. In systems with sufficiently great combinatorial complexity, generating the full reaction network becomes impossible. NFSim allows the assembly process to be simulated using a pool of interacting molecules, which are represented computationally as ‘agents’. Changes in the state and connectivity of the agents are determined by the rules, and quantities of relevant molecular species are tracked during simulation.