Kinetic Modeling of Biological Systems

Haluk Resat[1], **Linda Petzold**[2], and **Michel F. Pettigrew**[3]

[1] Pacific Northwest National Laboratory, P.O. Box 999, MS K7-90, Richland, WA 99352

[2] Department of Computer Science, University of California Santa Barbara, Santa Barbara, CA 93106

[3] Pettigrew Consulting, 2223 93rd Ave CT E, Edgewood, WA 98371

Abstract

The dynamics of how the constituent components of a natural system interact defines the spatio-temporal response of the system to stimuli. Modeling the kinetics of the processes that represent a biophysical system has long been pursued with the aim of improving our understanding of the studied system. Due to the unique properties of biological systems, in addition to the usual difficulties faced in modeling the dynamics of physical or chemical systems, biological simulations encounter difficulties that result from intrinsic multiscale and stochastic nature of the biological processes.

This chapter discusses the implications for simulation of models involving interacting species with very low copy numbers, which often occur in biological systems and give rise to significant relative fluctuations. The conditions necessitating the use of stochastic kinetic simulation methods and the mathematical foundations of the stochastic simulation algorithms are presented. How the well-organized structural hierarchies often seen in biological systems can lead to multiscale problems, and possible ways to address the encountered computational difficulties are discussed. We present the details of the existing kinetic simulation methods, and discuss their strengths and shortcomings. A list of the publicly available kinetic simulation tools and our reflections for future prospects are also provided.

Keywords

Biological network; kinetic simulation; stochastic simulation algorithm; tau-leaping; hybrid kinetic model; simulation software

1. Background and Introduction

Reactions occurring among a set of reactants define a kinetic reaction network. Traditionally, the set of reactions is described by nonlinear ordinary or partial differential equations, solutions of which provide insights into the dynamics of the studied processes. The size of such networks can vary considerably. In some instances, hundreds of thousands of reactions may be needed to describe the complex, coupled phenomena. Although kinetic modeling and the underlying fundamental mathematics are used in vastly different fields of science and engineering, here we will concentrate on biological kinetic networks which model cellular activities as a set of reaction processes. Biological networks are intrinsically very rich in character; feedback loops are common, and bifurcations and oscillations can lead to interesting dynamical behavior. Although there are many commonalities, two aspects, in particular, distinguish biological networks from networks found in other fields: Firstly, in biological systems, copy numbers of

many species are very low, which can give rise to significant relative fluctuations. Secondly, biological systems tend to have well-organized structural hierarchies.

The traditional way of modeling the time evolution of the molecular populations in a reacting system is to use a set of coupled, first order, ordinary differential equations (ODEs) called the *reaction rate equations* (RRE). The RRE are usually inferred from phenomenological arguments under the assumption that the system is *well-stirred* or *spatially homogeneous*; that is, no persistent correlations develop among the positions of the molecules. The RRE is nonlinear in the presence of bimolecular reactions, which is almost always the case. Owing to the possibility of widely differing reaction rates, the RRE can often exhibit vastly different time scales, a condition known as *dynamical stiffness*. Sometimes, under certain simplifying assumptions such as quasi steady-state or partial equilibrium, the RRE reduce to a system of differential-algebraic equations (DAEs).

However, a deterministic approach alone may not be sufficient for cellular systems when the underlying problem is inherently stochastic. Fluctuations in the concentrations of biomolecules can be quite large and this may require special treatment in modeling the biological networks (1–4). This is particularly true for molecular species involved in the transcriptional regulation of gene expression. In certain such cases, as the small copy number limit is reached, differential equation-based models break down since the assumption of continuous molecular concentrations is no longer valid. In such cases a more general approach, based on stochastic methods in which natural fluctuations are treated in a direct and explicit manner, has been advocated for the quantitative analysis of biological networks (5–7).

A simple argument demonstrates why fluctuations can be considerable in cellular systems. The typical diameter of a microbial cell is a few micrometers (μm), which corresponds to a cellular volume on the order of ~10 femtoliter (fl). In this volume, the concentration of 1 molecule is roughly equal to 160 picomolar (pM). This discrete nature is, of course, more severe when the molecules are confined to even smaller substructures, which is typically the case in eukaryotic cells. Thus, given that typical binding affinities of the biomolecules are often in the pM to μM range, and ignoring whether the thermodynamics are still valid at this limit, even very small fluctuations in the molecule copy numbers can cause a regime change in biomolecular interactions. Therefore, fluctuations resulting from a few discrete reactions can significantly affect the dynamical patterns in cellular systems. In a series of papers Gillespie showed the equivalency of deterministic and stochastic formalisms in the limit as the populations of all constituent chemical species tend to infinity, and discussed the importance of stochastic fluctuations when the system sizes are small (1,2). We note that stochastic differential equations are often used in physics, but they are usually obtained by starting with an assumed ordinary differential equation and then adding a "noise" term that has been carefully crafted to give a preconceived outcome. Although such approaches are computationally less expensive, they do not capture the true nature of the intrinsic fluctuations. Therefore, the use of *discrete stochastic approaches* is often more suitable in studying biological systems.

Although they provide a more realistic modeling frame, discrete stochastic simulations are often too slow and computationally expensive. Building upon earlier fundamental work (1,2, 8), development of new mathematical methods and numerical algorithms have been pursued by many groups to overcome some of the computational bottlenecks (5,6,9–23). Development of approximate algorithms was also pursued. For example, the probability weighted dynamic Monte Carlo method (24) has been shown to enable the simulation of network models with tens of thousands of reactions and hundreds of compartments using desktop computers (25), and various tau-leaping algorithms, which allow the use of larger time steps with tolerable inexactness (26–32), can also lead to significant speed ups in the stochastic simulations.

The spatial organization and structural hierarchies that often occur in biological systems provide an additional degree of complexity. Information transfer and biochemical activity can depend on location in cellular systems. Indeed, spatial organization becomes particularly important when the kinetics of the interaction networks depend on the local environmental factors, i.e., when they depend on local conditions and the environment or when material transport is required to convey the information. A good example to this is syntrophic (mutually interdependent) microbial systems, in which organisms in the community depend on each other for metabolite use and production. Here the metabolic activity necessary for cellular growth and survival can depend on the substrate availability and favorable environmental factors that are jointly established by multiple organisms. The relative concentration of the organisms and their geometrical arrangement can lead to heterogeneous distributions reflecting local variations in the microbial composition. Thus, the dynamical behavior of microbes in syntrophic systems relies on the combination of highly localized molecular events (i.e., metabolism) and macroscopic transport of products over a much larger spatial scale (i.e., between the cells of syntroph organisms). An analogous situation arises in mammalian systems, where eukaryotic cells have compartments with specific functionalities. These subcellular compartments exchange material in a concerted fashion as part of fundamental cellular processes such as synthesis, degradation, and receptor signaling. At the same time, the milieu of the cell within its tissue group is a consequence of the tissue itself. Thus, compartmentalized reactions, though many spatial scales smaller than the tissue in which they are found, simultaneously drive and respond to changes in the tissue. This hierarchy of compartmentalized activity represents a key challenge in multiscale modeling.

Particle transport dynamics can also play a central role in transferring information in many biological problems (33). This is particularly true when messenger substrate molecules or interacting proteins are transported over distances many orders of magnitude larger than their size. For example, energy metabolism of many bacteria requires cytochrome shuffling as part of the electron transfer mechanism. Oxygen availability and penetration can be very important in bacterial growth, both in batch reactors and in biofilms. Plant ecology too involves location specific activities. Most plants allow their root nodule cells to be infected by microbes or fungi for arbuscule development, which is used to obtain nutrients (particularly phosphorus) from the soil. At the same time, plants use their leaves for the necessary carbon production, some of which is also transferred to the microbial/fungal mycorrhizal symbiotic (intimate interdependent) partner. Material transport also plays a central role in determining cellular response in higher organisms, for which stress-induced Ca++/cAMP release in mammalian systems would be a good example (33).

Recent progress in experimental techniques, particularly fluorescent labeling and high-resolution microscopy, is beginning to make it possible to collect biological information about local dynamical activities (see, for ex., Refs. (34–38)). Models that represent dynamics using spatially averaged quantities cannot capture effects due to local inhomogeneities. For the types of problems described above, computational biology studies aiming to complement the experimental efforts require the development and analysis of spatially resolved multiscale models. The standard chemical master equation (CME) and stochastic simulation algorithm (SSA) assume that the system is well-stirred or spatially homogeneous (1,4), which for some problems might be an enormous simplification. Currently, the strategy most likely to work appears to be the *locally homogeneous* approach, which dates back to at least the mid-1970s (39,40). In this approach, the system volume is subdivided into K subvolumes that are small enough for each to be considered spatially homogeneous. Each subvolume forms a *reactor compartment*, and the compartments are linked at the higher whole-system scale. Using this framework, whenever a chemical reaction occurs in the system we now know where it occurred, at least to within the resolution of the subdivisions. However, this spatial resolution comes at a considerable cost in computational complexity. The dimensionality of the state space and the

number of chemical reaction channels have been increased by a factor of K, the number of subvolumes. Moreover, since a large set of reactions describing *the diffusive transfer between subvolumes* is needed to complete the model, the increase in computational complexity is in fact even greater.

From a purely mathematical point of view, the locally homogeneous approach to a spatially inhomogeneous system is essentially the same as for a spatially homogeneous system, in that one winds up with a CME and an SSA of the same "jump Markov" form (4,41,42). However, the resulting network model has a much higher dimensional state space and consists of a much larger set of reaction channels. The challenge of determining how to divide the system into subvolumes efficiently and without sacrificing the details of the problem must be overcome before this model can be routinely implemented. Another challenge is the development of algorithms that scale efficiently with the model dimension as it would be necessary to handle the complexity resulting from the steep increase in the network sizes. For large systems, the conspicuous absence of algorithms and software capable of taking advantage of available computing resources indicates a need for the development of hybrid algorithms in which the problem is separated into deterministic and stochastic parts to increase computational efficiency without sacrificing scientific veracity.

While both deterministic and stochastic formulations are currently used in the kinetic modeling of biological systems, as has been mentioned above, discrete stochastic methods are more suitable for biological systems. With this in mind, and because the numerical simulation of ordinary differential equation models is a well-established area with excellent and available software, in the remainder of this chapter we concentrate on stochastic simulation methods. We first summarize the fundamental steps and the mathematical foundations for kinetic simulation of reaction networks. We follow that with a discussion of shortcomings and size scaling properties of existing stochastic simulation algorithms. We then conclude by briefly highlighting the future research trends from our perspective.

2. SIMULATION METHODS

2.1 General Description

Kinetic simulation of biochemical systems involves the following basic steps:

1. Identification of the kinetic problem: This step involves the determination of the input and output variables as well as the intermediates. Key species and physical properties to be simulated are determined and tabulated.

2. Model formulation: Although that does not have to be the case, most studies employ the *reaction rate equations* (RRE) to model biochemical systems. In the RRE, one simply defines the changes in the concentrations (or equivalently the number of molecules) as a function of time and location. We note that constraints can be embedded in RRE using a Lagrange undetermined multipliers type formalism.

3. Choosing a method: At this step a decision needs to be made as to whether to adopt a deterministic or stochastic formulation. Although it is still relatively uncommon, a hybrid approach which moves between deterministic and stochastic regimes is another possibility. In the deterministic approach, a reasonable time step is chosen adaptively, based on the estimated local error according to a differential equation model. In contrast, discrete stochastic simulations model each individual reaction event. Using the probabilities of the reactions (called the *reaction propensities*), one determines the occurrence statistics of the involved reactions. This is generally done in one of two ways: In next reaction type methods, the sequence of the reactions are chosen according to the reaction propensities and the reaction times are computed to

determine the elapsed time. In lumped reaction approaches, one can choose the size of the elapsed time step and compute how many times the involved reactions occur according to their probabilities.

4. Simulation: Integrating the RRE deterministically involves choosing an appropriate algorithm from a wide selection of sophisticated numerical methods for systems of ODEs and DAEs. Highly efficient and reliable software is readily available, although one must usually determine whether or not the problem is stiff and then choose the software accordingly. Roughly speaking, an ODE or DAE system is stiff if it involves a wide range of time scales and the fastest of those scales correspond to stable processes. In chemically reacting systems, the wide range of time scales can come from some reaction rates being several orders of magnitude or more greater than others. A solver which is designed for non-stiff systems will run very slowly (because it needs to choose time steps on the scale of the fastest process in the system to maintain stability for its explicit formulas), if it is applied to a stiff problem. The cure for stiffness is to approximate the differential equation using implicit methods. The software that is available for the accurate solution of stiff ODE and DAE systems always makes use of implicit methods. For more information on ODE and DAE methods and software, see Ascher and Petzold (43).

Stochastic simulation algorithms are not as advanced as the deterministic integrators and they do not scale well with the model size and complexity. Even though recent algorithmic improvements have led to impressive gains, stochastic simulations are still too slow for most realistic systems. However, as discussed in the previous section, fluctuations can be large in biological systems and, more importantly, stochastic variations may have implications for biological functions and responses. Therefore, when feasible, stochastic simulation methods should be preferred. This is particularly true for processes that involve small number of regulatory molecules with large variations in their expression levels.

5. Trajectory analysis: Time courses obtained during the simulations are catalogued and combined to derive the statistical distributions of the desired output/prediction quantities, such as the concentrations of experimentally monitored species and the resulting material/mass flow in the system or the probability distribution of the reactions. Similarly, from simulation results for different model parameter sets, one can study the sensitivity to parameter variations.

The following simple example illustrates steps (1–3). Assume that we are studying a receptor (R) system that binds a ligand L. The ligand bound receptors (RL) binds an effector/adaptor/scaffold molecule E that helps with signal transduction. Receptors that are in complex with the effector molecules are labeled as RLE. For this case, the kinetic system consists of the following two reversible reactions L+R RL and RL+E RLE. Forward and reverse reactions for these two reactions are second and first order reactions, respectively. For simplicity, let us use a mass-action formalism where the rates for the first (A→C) and second (A+B→C) order reactions are represented as k[A] and k[A][B], respectively, where [X] stands for the concentration (or the number of molecules with the proper volume conversion) of species X.

In this example, the species list would be L, R, RL, E, and RLE. The reaction list will contain four reactions (i) L+R→RL, (ii) RL→L+R, (iii) RL+E→RLE, and (iv) RLE→RL+E. If RLE is the species of interest (signal transducing receptors), the output of the simulations would be [RLE], the concentration of the species RLE. This identification completes the step (1).

Development of the RRE follows a simple rule: Reactions that decrease or increase the concentration of a species appear as a source term in the rate equation for that species. For example, the reaction RL+E→RLE decreases [RL] and [E], and increases [RLE]. So the term

k$_3$[RL][E] is included as a negative contribution in the rate equations for d[RL]/dt and d[E]/dt, and as a positive contribution in the d[RLE]/dt rate equation. Using this logic the RRE equations for the above example would be

$$d[L]/dt = -k_1[L][R] + k_2[RL]$$
$$d[R]/dt = -k_1[L][R] + k_2[RL]$$
$$d[RL]/dt = +k_1[L][R] - k_2[RL] - k_3[RL][E] + k_4[RLE]$$
$$d[E]/dt = -k_3[RL][E] + k_4[RLE]$$
$$d[RLE]/dt = +k_3[RL][E] - k_4[RLE]$$

Once the model is defined as above, preparation for the simulation is slightly different depending on the method to be used. If a deterministic approach is chosen, then the RRE is investigated in terms of the stiffness of the equations, and an appropriate integrator is chosen. We note that for more general formalisms, singularity of the Jacobian is another point of concern but that subject is outside the scope of this chapter. If a stochastic simulation approach is chosen, then the RRE is converted into a reaction table (Table 1) form where propensities are associated with each reaction.

In the stochastic approach, the stoichiometry matrix defines the change in the species concentrations when the corresponding reaction occurs. Entries for the reactants and the products appear as negative and positive, respectively. With this information at hand, at every step of the stochastic simulation one can update the system configuration $\mathbf{x}$(t) according to the occurrence of reactions as $\mathbf{x} \rightarrow \mathbf{x} + v \cdot \omega$ using the information about how many times the reactions occur ω.

Step (4), execution of the simulations, is the key step in the kinetic modeling studies. The numerical solution of ODEs or DAEs is a well established area with widely available software (43). Thus, we will focus on stochastic simulation methods in the remainder of this chapter.

2.2 Stochastic Simulation Algorithms - Fundamentals

The Stochastic Simulation Algorithm (SSA) developed by Gillespie (1–4) and its variants are the basis for stochastic kinetic simulation of biological systems. In SSA it is assumed that successive *reactive* molecular collisions in the system are separated by many *non-reactive* collisions, which effectively wipe out all positional correlations among the molecules. Therefore, SSA is suitable for the simulation of well stirred systems.

Defining the *state* of the system by the vector $\mathbf{x}$(t) whose components $x_i(t)$ (i=1 to N for N species) donate the number of molecules of species S_i in the system at time *t*, one can establish from kinetic theory the following result: For each elemental chemical reaction channel R_j (j =1,...,M), there exists a function a_j such that, if the system is in state $\mathbf{x}=(x_1,\ldots, x_N)$ at time *t*, then the *probability* that reaction R_j will occur somewhere inside the system in the next infinitesimal time *dt* is $a_j(\mathbf{x})$. This can be regarded as the fundamental premise of stochastic chemical kinetics. The function a_j is called the *propensity function* of reaction channel R_j. Each reaction channel R_j is completely characterized by its propensity function together with its *state change vector* $v_\mathbf{j}$ which represents the change in the population of the molecular species caused by one R_j event, *i.e.*, the stoichiometric coefficients of the reactions. Thus, if the system is in state $\mathbf{x}$ and one R_j reaction occurs, the system jumps to state $\mathbf{x} + v_\mathbf{j}$. This is mathematically known as a *jump Markov process* and a great deal is known about such processes (8,41).

The stochastic evolution of the state vector $\mathbf{x}(t)$ is specified by the function $P(\mathbf{x},t|\mathbf{x_0},t_0)$, defined as the *probability* that $\mathbf{x}(t)$ will equal $\mathbf{x}$ given that $\mathbf{x}(t_0)= \mathbf{x}_0$ for $t_0 \leq t$. It can be proven that this function obeys the following time-evolution equation (4,44):

$$\frac{\partial P(\mathbf{x}, t|x_0, t_0)}{\partial t} = \sum_{j=1}^{M} [a_j(\mathbf{x} - \nu_j)P(\mathbf{x} - \nu_j, t|x_0, t_0) - a_j(\mathbf{x})P(\mathbf{x}, t|x_0, t_0)]$$

(1)

This elegant equation is known as the *chemical master equation* (CME). Unfortunately, it is of such a large dimension (it includes a differential equation corresponding to every possible state of the system) that its solution is almost always intractable, both analytically and numerically.

An alternate way of analyzing the behavior of $\mathbf{x}(t)$ is to construct its unbiased realizations using the SSA (1,2). The steps of SSA basically are:

i. Initialize the system's state $\mathbf{x}=\mathbf{x}_0$ at time $t=t_0$;

ii. Evaluate the reaction propensities $a_j(\mathbf{x})$ and their sum $a_0(\mathbf{x})$;

iii. Draw two random numbers r_1 and r_2 from a uniform distribution in the unit-interval, and compute the time interval between the reactions $\tau = \ln(r_1)/a_0(\mathbf{x})$ and determine which reaction type R_j occurs next by finding the smallest integer j that satisfies

$$\sum_{j'=1}^{j} a_{j'}(\mathbf{x}) > r_2\, a_0(\mathbf{x}) \quad ;$$

iv. Record and update the changes in the system $t \rightarrow t+\tau$ and $\mathbf{x} \rightarrow \mathbf{x}+\mathbf{v_j}$; and

v. Return to step (ii) or end the simulation.

The SSA advances the system in time from one reaction event to the next, and it does so in a way that is rigorously equivalent to the CME. The SSA is therefore an *exact* method for simulating the time evolution of a spatially homogeneous chemical system. Because SSA processes reaction events serially, i.e., one after another, it is often too slow to be practical when applied to real systems of interest. An extremely large molecular population of even one reactant species, and/or a very fast reaction, will usually cause the time increments τ to be extremely small.

2.3 Acceleration of Discrete Stochastic Simulation

Several lines of research have been pursued in attempts to improve upon the exact SSA. These include more efficient implementations of the SSA and the development of approximate SSAs with improved numerical efficiencies.

The efficiency of the algorithms can strongly depend on the formulation and numerical implementation and these are key aspects to consider when developing simulation software. Gibson and Bruck used an efficient implementation strategy for the exact SSA (10) to remove a key limitation, which restricted the practical use of the first reaction method of Gillespie (1), by not requiring the generation of N_r new random uniform deviates following a reaction event where N_r is the number of active reactions. Rather than a single waiting time between events, the method stores the absolute firing time for each reaction in an indexed priority binary heap queue with times increasing with increasing depth in the heap. Assuming a static reaction network topology, a reaction dependency graph is created initially. Following an event, which is identified by the element at the top of the heap, a single random uniform deviate is generated, reaction propensities are efficiently updated using the dependency graph, and any absolute

firing time requiring updating is adjusted through a linear transformation. Lastly, the heap queue is reordered. Gibson and Bruck have shown that the algorithm has a time complexity of order $O(N_r + N_E \log N_r)$ where N_E is the number of events and generally requires only a single random uniform deviate per event (45). While the next reaction method can be efficient for large, very sparse networks, significant overhead may incur in maintaining the indexed priority queue. This has led Cao et al. to conclude that, for most kinetic systems, the most efficient formulation of the SSA is an optimized direct Gillespie method (46).

One way to speed up the next reaction method type algorithms is to allow multiple reactions to take place in longer time steps, which is the underlying idea behind the probability-weighted dynamic Monte Carlo (PW-DMC) method (24) and the tau-leaping based algorithms (28–32, 47). The basic premise behind the PW-DMC method is that reactions with large propensities dominate the stochastic simulation. By attaching weights that are updated as the simulation progresses, the PW-DMC approach attempts to equalize the propensities of the reactions constituting the network model. In this process, the large propensities are scaled down with an associated factor, which corresponds to how many times the involved reaction occurs when selected. It was shown that the PW-DMC algorithm solves the multiple time scale problem to a large extent (24). Efficient statistical sampling obtained with the PW-DMC algorithm made the simulation of an integrated stochastic model of EGFR signaling network that contains tens of thousands of reactions possible using desktop computers in reasonable wall clock periods (25). The success of PW-DMC results from its effectiveness in bundling the occurring reactions, and in this aspect, it is conceptually similar to the approach pursued in tau-leaping methods. The main differences between the PW-DMC and tau-leaping methods are that in the PW-DMC the bundling of the reactions is done for each reaction type (i.e., some reactions can be selected to be left out if desired) and the integration time does not have to be set in advance, but the tolerance in the fluctuation levels for the species need to be predefined. These features make PW-DMC suitable to deal with the reaction networks that can be partitioned into subclasses according to their spatio-temporal properties where the allowed fluctuation levels can be dictated on a per reaction basis. It was shown (24) that, while the fluctuation variations may be affected, the PW-DMC conserves the mean quantities. In this respect, PW-DMC is a weak order (i.e., not a strong order) stochastic method. The fluctuations in the computed quantities depend on the achieved speed-up, which can be several orders of magnitude without any significant effect on the computed mean quantities (24).

In the tau-leaping (τL) methods (47), advances in the system configuration are determined based on a *preselected* time step τ during which more than one reaction event may occur. Tau-leaping is based on the same premises that underlie the CME and the SSA: If τ is chosen small enough to satisfy the *leap condition*, which says that none of the propensity functions may change noticeably during τ, then given $\mathbf{x}(t)=\mathbf{x}$, the system state at time $t+\tau$ can be *approximately* computed as

$$x(t+\tau) = x + \sum_{j=1}^{M} P_j(a_j(x), \tau) v_j.$$

(2)

The P_j are statistically independent Poisson random variables; they physically represent the number of times each reaction channel fires in time τ. The Poisson random variable $P(a, \tau)$ is in fact defined as the number of events that will occur in time τ, given that $a\, dt$ is the probability of an event occurring in the next infinitesimal time dt. We emphasize that τ must be chosen small enough that none of the propensity functions change significantly during the leap. Algorithms have been developed for estimating the largest value of τ that will satisfy that condition for a prescribed level of simulation accuracy (26,28).

While the τL methods show significant gains in efficiency, the unbounded Poisson distribution can lead to nonphysical system configurations in which negative species populations arise when the expected number of events is large enough to consume the reactant populations. Cao et al. have shown how to avoid negative populations in explicit Poisson τL (PτL) methods (48). The non-negative PτL algorithm is based on the fact that negative populations typically arise from multiple firings of reactions that are only a few firings away from consuming all the molecules of one of their reactants. To focus on those reaction channels, the modified PτL algorithm introduces a control parameter n_c, a positive integer that is usually set somewhere between 10 and 20. Any reaction channel that is currently within n_c firings of exhausting one of its reactants is then classified as a critical reaction. The algorithm chooses τ in such a way that no more than one firing of all the critical reactions can occur during the leap. Essentially, the algorithm simulates the critical reactions using SSA, and the remaining non-critical reactions using τ-leaping (48).

An alternative approach is based on choosing the reaction numbers using binomially distributed random variables that have the appropriate finite sampling ranges. If a short time interval τ_Σ is considered, for each reaction R_j, the Poisson distribution is replaced with a binomial distribution $B(p_j, \omega_j^*)$ with the same mean $a_j\tau$. Here ω_j^* is an upper bound (or limit) on the reaction number and p_j is the probability that Rj will fire in the interval τ_Σ/ω_j^*. In the resulting binomial tau-leaping (BτL) method, the leap size is then chosen as the smaller of the times t_j computed from the constraints $p_j=a_jt_j/\omega_j^* \leq 1$, j=1:M and the upper bound τ_ε. Once the tau step has been determined, each network reaction number ω_j is generated by sampling the corresponding binomial distribution.

Although the binomial distribution has the desirable property of having a finite sampling range, this in itself does not completely resolve the non-negativity problem. In networks with multiple channel reactant dependencies, the interdependence arising from species participating in multiple reactions may still generate negative species populations. Tian and Burrage have shown how the BτL completely resolves the non-negativity problem when there is at most a two-reactant dependence in a reaction network (31). In the BτL method of Chatterjee et al. (29,30) the issue of multiple-channel reactant dependencies is dealt with by assigning to each reaction a binomial random variable for the number of reaction events during a step and by modifying reaction number limits through tracking currently available population sizes as reactions fire during that step. Yet, as critically noted by Chatterjee et al., in the latter approach selected reaction numbers may depend on the order in which reactions are processed (29,30).

The multinomial tau-leaping (MτL) method of Pettigrew and Resat efficiently extends the BτL method to networks with arbitrary multiple-channel reactant dependencies (32). Improvements were achieved by a combination of three factors: First, tau-leaping steps are determined simply and efficiently using a-priori information and Poisson distribution based estimates of expectation values for reaction numbers. Second, networks are partitioned into closed groups of reactions and corresponding reactants in which no group reactant set is found in any other group. Third, product formation is factored into upper bound estimation of the number of times a particular reaction occurs. Together, these features allow for larger time leaping steps while the numbers of reactions occurring simultaneously in a multi-channel manner are estimated accurately using a multinomial distribution (32). Partitioning of the reaction network into groups with closure can be done with a graph theoretic algorithm based on the Dulmage-Mendelsohn decomposition of the stoichiometric matrix representing the network. Two methods for selecting an upper limit Ω^* on the system reaction number was advocated, the rate limiting reactant and the rejection methods (32). The former is a conservative approach that guarantees that the chosen reactions do not lead to negative populations after the system's configuration is updated according to the occurring reactions. In contrast, the selection process in the rejection method only guarantees that the system

configuration after the tau-leaping step is non-negative. It however does not ensure that all possible multi reaction paths that take the system to that final configuration involve only physically feasible intermediate stages. It was argued that, since one can only detect the relationship between the starting and end points of a move during the simulation, ensuring the state of the system to always correspond to physically feasible configurations at the observation points (i.e., starting and end points of the moves in a simulation run) is the most crucial aspect in obtaining accurate results (32) . Therefore, allowing a simulation trajectory to temporarily cross the solution space boundary would introduce insignificant inaccuracies because statistics for intermediate states have only an indirect impact. In the MτL method, tau-leaping step is then chosen as the smaller of the times t computed from the constraint $p = t\Sigma/\Omega^* \leq 1$ and the upper bound τ_Σ where Σ is the system propensity. The actual system reaction number Ω is then sampled from a binomial distribution as discussed above, and the reaction numbers ω_j, j=1:M −1, are drawn from a multinomial distribution with corresponding probabilities $\alpha_j= a_j/\Sigma$ and $\omega_M=\Omega−\omega_1−\omega_2− …−\omega_{M-1}$. Implemented advances over the BτL method were shown to lead to significant speed ups in the simulations of biologically relevant models (32). It was also shown that MτL is less sensitive to the error parameter of the τL methods, which is a desirable aspect of the kinetic simulation algorithms.

2.4 Thermodynamic Limit

It is well known that the *Poisson* random variable P(a,τ), which has mean and variance $a\tau$, can be approximated by a *normal* random variable with the same mean and variance when $a\tau \gg 1$. Using this fact, it can be shown that if τsatisfies not only the leap condition but also the conditions $a_j(\mathbf{x})\tau \gg 1$ for *all j*, which physically means that during time τ every reaction channel can be expected to fire many more times than once, then the basic tau-leaping Eq. (2) further approximates to (49)

$$X(t+\tau)=x+\sum_{j=1}^{M} \nu_j a_j(x)\, \tau+\sum_{j=1}^{M} \nu_j \sqrt{a_j(x)\tau}N_j(0,1).$$

(3)

Here the $N_j(0,1)$ are statistically independent normal random variables with mean 0 and variance 1. This modified update formula is computationally faster than the tau-leaping formula (2) which it approximates, because samples of the normal random variable are much easier to generate than samples of the Poisson random variable. Also, since by hypothesis every reaction channel fires many more times than once during each time step τ, simulations using the updating Eq. (3) will be *very* much faster than the SSA.

If the arguments leading to Eq. (3) are repeated with τ replaced by a *macroscopic infinitesimal dt*, which by definition is *small* enough that no propensity function changes significantly during *dt* yet *large* enough that every reaction channel fires many more times than once during *dt*, then one can deduce the equation

$$\frac{dx(t)}{dt}=\sum_{j=1}^{M} \nu_j\, a_j(x(t))+\sum_{j=1}^{M} \nu_j \sqrt{a_j(x(t))}\Gamma_j(t).$$

(4)

where the gammas are statistically independent *Gaussian white noise* processes (49–51).This equation is called the *chemical Langevin equation* (CLE), and it is an example of what is known to mathematicians as a *stochastic differential equation*. It is entirely equivalent to the *Langevin leaping formula* given by Eq. (3).

We note that stochastic differential equations are often used in physics, but they are usually obtained by starting with an assumed ordinary differential equation and then adding a "noise" term that has been carefully crafted to give some preconceived outcome. So it is noteworthy that the noise term in Eq. (4), namely the second summation there, has actually been *derived* from the same premises that underlie the CME and the SSA using physically transparent assumptions and approximations. Use of the Langevin formulas (3) and (4) is typically justified whenever all reactant species are present in sufficiently large numbers.

At the *thermodynamic limit*, molecular populations of all the reactants *and* the system volume become infinitely large, while the concentrations of all species stay constant. It can be proven that, in this limit, all propensity functions scale *linearly* with the system size. Thus, the deterministic terms on the right sides of Eqs. (3) and (4) grow in proportion to the system size, while the fluctuating terms grow in proportion to the *square root* of the system size. This is another way of showing that for large systems "relative fluctuations in the species populations scale as the inverse square root of the system size", a central rule of thermodynamics. In the *full thermodynamic limit*, the foregoing scaling behavior implies that the fluctuating terms in Eqs. (3) and (4) become negligibly small compared to the deterministic terms, and the CLE reduces to

$$\frac{d\mathbf{x}(t)}{dt} = \sum_{j=1}^{M} v_j a_j(\mathbf{x}(t)).$$

(5)

This is none other than the traditional *reaction rate equation* (RRE), which is more commonly used in terms of the concentration vector. Note that the RRE has not been assumed here; instead, it has been *derived* from the same premises that underlie the CME and the SSA, using a series of physically understandable approximations (49,50). Thus, the discrete, stochastic CME/SSA and the continuous, deterministic RRE are actually connected across the end of the representation spectrum. The challenge, of course, is to develop both the software infrastructure and the theory necessary to make reliable adaptive decisions so that we can accurately and efficiently simulate complex reaction systems based on detecting a discrete or continuous regime and using an optimal blend of stochastic and deterministic methods.

2.5 Scaling Properties With Respect to the Network Size

In addition to the multiscale nature of the involved kinetic processes, computational efficiency of stochastic simulation algorithms also depend on several measures of the size of the investigated network model, such as the overall complex population, the total number of reactions, and the average number of nodal interactions or connectivity in a network. Understanding how algorithms scale with size and nodal connectivity is an important consideration in the design of numerical experiments for dynamic biological systems. As the complexity of the systems studied in biological systems increases, such scaling issues become a serious point of concern.

To study how stochastic kinetic simulation algorithms scale with network size, construction of scalable reaction models that can be used in benchmark studies has been advocated (42). Pettigrew and Resat used two scalable reaction models to compare the efficiency of the exact Gillespie algorithm as implemented using the popular Gibson-Bruck method (GGB) (10) with the inexact random substrate method of Firth and Bray (FB) (52). Foundations of these two algorithms are quite distinct from each other and the latter algorithm has been stated to be an efficient method when the total complex population is small and the number of multistate interactions large.

The first investigated scalable model, the Size Scalable Model (SSM), is a four major-compartment model that is typical for a signal transduction network in eukaryotic cells (42). The SSM involves 2 unimolecular and 5 bimolecular reversible reactions with mass transfer between compartments for 8 complex types including ligands, adapter complexes, and receptors (42). Upon phosphorylation the receptor can become active in signal transduction. A sub-compartmentalization process, which is equivalent to forming smaller locally homogeneous regions, allows for the creation of refined models where the size of the model increases in proportion to the number of created subcompartments while maintaining the connectivity (i.e., the topology) between the species. For SSM, it was found that the GGB algorithm, whose numerical efficiency scales with the logarithm of the total number of active reactions, performs significantly better in all cases including those characterized by a very small number (~100) of complex types, each with a very small population (42).

The second scalable model, the Variable Connectivity Model (VCM), is a single compartment signal transduction network involving 2 unimolecular and 2 bimolecular reversible reactions for 3 complex types, ligand, receptor, and ligand receptor complexes (42). The receptors in this model have multiple phosphorylation sites. The size and the node connectivity between the species increase in proportion to the number of phosphorylation sites N. For the VCM, it was shown that the FB random substrate method, whose efficiency scales with the square of the total complex population, can outperform the GGB algorithm when the node connectivity is sufficiently high and the total complex population sufficiently low. Performance of the FB and GGB algorithms were determined as a function of the number of phosphorylation sites for two distinct population sets. In the first set the total complex population was initially set at 192 complexes (128 ligands and 64 receptors), while for the second set the total complex population was 10 times higher with the 2:1 ratio of ligands to receptors unchanged. It was found that, for the smaller population set the FB algorithm becomes more efficient than the GGB when the number of phosphorylation sites N reaches 6, whereas in the larger population set the GGB is clearly superior for small N up to a crossover point that occur around $N{\sim}11$. Thus, it was concluded that with the exception of special cases, GGB is a more efficient method than the FB (42).

This simple study using two sample realistic networks clearly illustrated that the most efficient algorithm can depend on the problem type and the used simulation method should be chosen after a careful study of the topology and multiscale properties of the investigated network.

3. Simulation Tools

Table 14.2 lists and briefly describes some of the existing software tools that may be used in the kinetic simulations of biological systems. In this section we list and briefly describe some of the existing software tools that may be used in the kinetic simulations of biological systems. We would like to note that the list is in no way complete and it only represents a small percentage of the available tools.

NAME	Brief Description
BIONETGEN	Automatic generation of physicochemical models of biological systems with combinatorial complexity such as cellular signaling from user-specified rules for biomolecular interactions at the protein domain level, integrated with tools for reaction network simulation and analysis (53). http://cellsignaling.lanl.gov/bionetgen/
BIO-SPICE	Open source software for intra and inter-cellular modeling and simulation of spatio-temporal processes, pathways and interaction-network tools for data analysis (MIAMESpice), model composition and visualization (BioSenS, Charon, JDesigner, Sal), mathematical tools including ordinary and partial differential and stochastic equation-based methods and algorithms for the simulation of reaction and diffusion-reaction networks (BioNetS Simpathica) (54,55). https://biospice.org/index.php

CELLDESIGNER   Gene regulatory network modeling, visual representation of biochemical semantics with SBML and database support, integration with Matlab or Mathematica ODE Solvers and direct access to SBW SBML compliant simulators such as Jarnac (56,57).
http://celldesigner.org/index.html

CELLERATOR   Open source Mathematica based package for simulation and analysis of signal transduction networks in cells and multicellular tissues, automated equation generation with arrow-based reaction notation (58).
http://www-aig.jpl.nasa.gov/public/mls/cellerator/

COPASI   Stochastic and deterministic simulation of network pathways, steady state and metabolic control analysis including stability analysis, elementary and mass conservation analysis, optimization and parameter estimation and SBML support (59).
http://www.copasi.org/tiki-index.php

DIZZY   Stochastic and deterministic kinetic modeling of integrated large-scale genetic, metabolic, and signaling networks with domain compartmentalization, features include modular simulation framework, reusable modeling elements, complex kinetic rate laws, multi-step reaction processes, steady-state noise estimation (60).
http://magnet.systemsbiology.net/software/Dizzy

E-CELL   Object-oriented software for modeling, simulation, and analysis of large scale cellular networks, multi-algorithm multi-time scale simulation method with access to Gillespie-Gibson and Langevin stochastic algorithms, Euler and higher order adaptive methods for ordinary and algebraic differential equations, parallel and distributed computing capabilities and SBML support (61,62).
http://www.e-cell.org/software/e-cell-system

GRID CELLWARE   Grid-based modeling and simulation tool for biochemical pathways, integrated environment for diverse mathematical representations, parameter estimation using swarm algorithm and optimization, user-friendly graphical display and capability for large, complex models, stochastic algorithms such as Gillespie, Gibson and tau-leaping and deterministic algorithms based on ordinary differential equation solvers (63).
http://www.bii.a-star.edu.sg/research/sbg/cellware

MCELL   Monte Carlo simulator for investigating cellular physiology, model description language or MDL used to specify ligand creation, destruction, and release as well as chemical reactions involving effectors such as receptors or enzymes, ligand diffusion modeled with 3D random walks, wide variety of numerical and imaging options (64).
http://www.mcell.cnl.salk.edu and www.mcell.psc.edu

MESORD   Open source C++ software for stochastic simulation of kinetic reactions and diffusion based on next subvolume method (NSM) for structured geometries in three dimensions, SBML compatible, three-dimensional OpenGL simulation viewer (65).
http://mesord.sourceforge.net/

NWKSIM   Fortran 90/95 platform for large scale general kinetic reaction network simulation, Java swing graphical user interface, unique modeling features for computation of evolving multiple compartments with trafficking and signal transduction processes, stochastic algorithms include direct Gillespie, probability weighted dynamic Monte Carlo and weighted Gibson-Gillespie Bruck, binomial and multinomial tau-leaping algorithms (32,42).

SIGTRAN   Fortran 90/95 modeling and simulation platform for large scale reaction networks, Java swing graphical user interface, dual stochastic and deterministic (ODE/DAE) simulation modes, multistate macromolecule specification and simulation using the Firth-Bray algorithm, reaction-diffusion modeling with Fricke-Wendt algorithm, molecule tagging and tracking in signal transduction networks using a fully automated graph-theoretic algorithms for determination of unambiguous set of base species, SBML file support (42).

SIMBIOLOGY   Matlab toolkit for modeling, simulating, and analyzing biochemical pathways, graphical user interface with visual pathway expression, manual or SBML file input, stochastic or deterministic solvers, parameter estimation and sensitivity analysis, ensemble runs and post-run analysis tools with plotting.
http://www.mathworks.com/products/simbiology/description4.html

SMARTCELL   Object-oriented C++ platform for modeling and simulation of diffusion-reaction networks in whole-cell context with support for any cell geometry with different cell compartments and species localization, includes DNA transcription and translation, membrane diffusion and multistep reactions as well as cell growth, localization and diffusion modeling based on mesoscopic stochastic reaction algorithm (66).
http://smartcell.embl.de/introduction.html

STOCHKIT   Efficient C++ stochastic simulation framework for intracellular biochemical processes, stochastic algorithms includes Gillespie SSA and explicit, implicit and trapezoidal tau-leaping methods, Kolmogorov distance and histogram distance for quantifying difference quantification in statistical distribution shapes via Matlab, extensible to new stochastic and multiscale algorithms (67).
http://www.engineering.ucsb.edu/~cse/StochKit/StochKit.html

SBTOOLBOX   Matlab toolbox offering open and user extensible environment for prototyping new algorithms,and building applications for analysis and simulation of biological systems,

deterministic and stochastic simulation, network identification, parameter estimation and sensitivity analysis, bifurcation analysis, SBML model import (68). http://www.sbtoolbox.org/

VIRTUAL CELL Model creation and simulation of cell biological processes, associates biochemical and electrophysiological data for individual reactions with experimental microscopic image data describing subcellular locations, cell physiological events simulated within empirically derived geometries, reusable, updatable and accessible models, simulation data stored on the Virtual Cell database server and is easily exportable in variety of formats (69,70). http://www.vcell.org/

4. Future Prospects of Stochastic Simulation

In recent years there has been rapid progress in the analysis and development of algorithms for the accelerated discrete stochastic simulation of chemically reacting systems. Although these developments provide a good foundation, existing stochastic simulation methods are far from being computationally efficient to simulate detailed biological models. This is particularly true for the models that involve multiple time and spatial scales while still requiring high resolution for accurate representation. Thus, the development of efficient and adaptive multiscale algorithms and software for stochastic-deterministic simulations are urgently needed. This is particularly true for the simulation of spatially resolved models because immense recent progress in imaging is beginning to make the data about the inhomogeneous distribution of molecular species available. As the expense associated with spatially resolved models makes fully stochastic simulations nearly infeasible, efforts to develop hybrid approaches will become very important. Hybrid methods (22,71) have recently been proposed to simulate multiscale chemically reacting systems. These methods combine the traditional deterministic reaction rate equation, or alternatively the chemical Langevin equation, with the SSA. The idea is to split the system into two regimes: the continuous regime and the discrete regime. The RRE is used to simulate the fast reactions between species with large populations and the SSA is used for slow reactions or species with small populations. The conditions for the continuous regime are 1) the number of instances of each molecular species in a reaction in the continuous regime must be large relative to one, and 2) the number of reaction events of each reaction occurring within one time step of the numerical solver must be large relative to one (72). If either condition is not satisfied for a reaction channel, that reaction channel must be handled in the discrete regime.

The hybrid methods efficiently use the multiscale properties of the problem but there are still some fundamental unsolved problems. The first is the automatic partitioning of systems. This needs to be done very carefully, because both accuracy and efficiency depends on a proper partitioning. It will require a precise awareness of the assumptions under which algorithms and approximation available in the multiscale framework are valid, and the corresponding tests on the system to determine the partitioning. This is one reason why a proper theoretical foundation for each algorithm is so critical — to ensure that we are using it only to make approximations for which the corresponding assumptions are valid. Concerns about automatic step size selection continue to exist. Another difficult problem arises in the case when a reaction is fast but one of the corresponding reactants has a small population thus failing to satisfy the second requirement above for the deterministic regime. Such a reaction channel will still be handled by the SSA, resulting in a very slow simulation. New approaches and approximations, such as the slow-scale SSA (73) have the potential to overcome this limitation.

Another promising area of research is the application of tau-leaping based stochastic algorithms to problems involving reaction-diffusion processes. When the SSA algorithm is applied to reaction-diffusion processes the standard approach is to subdivide the heterogeneous system volume into a sufficiently large number of small-volume elements $V_0 \sim O(h^3)$ so that the assumption of a homogeneous, well-stirred system within each element is recovered. With diffusion considered as a first-order reaction in which molecules of a species are exchanged between neighboring elements at a rate $r_D \propto V_0 D/h^2$, kinetic reactions within an element occur

at a rate $r_C \propto V_0 \Sigma$. Here D and Σ are, respectively, the diffusion constant and the total system propensity. By choosing elements sufficiently small so that $r_C \ll r_D$ is true for all elements, the SSA may be applied although under these circumstances it is not usual to be confronted with a reaction network where the total number of processes and species is now several orders of magnitude larger than that found in a single element. Moreover, it has been observed in stochastic simulations with the SSA that reaction-diffusion processes are strongly dominated by diffusion events. The increase in network size and diffusion dominated events both contribute to a severe reduction in efficiency and it is highly likely that a number of more recent stochastic algorithms for reaction-diffusion suffer from the same problem. It is possible that τ-leaping methods may be well suited to overcoming this issue with the stochastic simulation of reaction-diffusion problems.

As stochastic simulation is used more and more often in studying the dynamics of biological systems, and as biological models become larger and more complicated, there is a need for the development and efficient implementation of simulation algorithms that can handle the increasing complexity. Thus, the issue of how an algorithm scales with the system size is crucial in assessing the limitations of the algorithm. Scaling concerns expose the serious need for establishing realistic scalable models that can be used in benchmark studies for algorithm efficiency comparison. Pettigrew and Resat (42) have introduced two such models. It was observed (42) that the scaling qualities of the algorithms investigated break down significantly when the network reaches a certain size. We suspect that this is a common occurrence for all of the existing algorithms, which unfortunately can impose severe limitations on the utility of the stochastic kinetic approaches. Roughly estimating, existing stochastic simulation algorithms can be used to investigate reaction networks containing $\sim 10^5$ reactions using a desktop machine. Utilization of high-performance resources may push this limit to 10^8. Considering that even the smallest microbial community or physiologically relevant eukaryotic network can easily contain $> 10^{12}$ reactions, it is clear that the scaling properties of the stochastic simulation algorithms need significant improvements before they can be widely used for larger problems. As discussed above, hybrid methods that reduce the computational complexity by treating the non-critical parts of the network using deterministic models may offer the most suitable solution to the scaling bottleneck in the near future.

We conclude by observing that while recent improvements in algorithm development, computational implementation, and hybrid approaches are indications of a promising future for the stochastic kinetic simulation of biological systems, there is still plenty of room for achieving significant advances.

References

1. Gillespie DT. A general method for numerically simulating stochastic time evolution of coupled chemical-reactions. Journal of Computational Physics 1976;22:403–434.

2. Gillespie DT. Exact stochastic simulation of coupled chemical-reactions. Journal of Physical Chemistry 1977;81:2340–2361.

3. Gillespie DT. Concerning validity of stochastic approach to chemical-kinetics. Journal of Statistical Physics 1977;16:311–318.

4. Gillespie DT. A rigorous derivation of the chemical master equation. Physica A 1992;188:404–425.

5. Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. Genetics 1998;149:1633–1648. [PubMed: 9691025]

6. McAdams HH, Arkin A. Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A 1997;94:814–819. [PubMed: 9023339]

7. McAdams HH, Arkin A. It's a noisy business! Genetic regulation at the nanomolar scale. Trends in Genetics 1999;15:65–69. [PubMed: 10098409]

8. Bortz AB, Kalos MH, Lebowitz JL. New algorithm for Monte-Carlo simulation of Ising spin systems. Journal of Computational Physics 1975;17:10–18.

9. Endy D, Brent R. Modelling cellular behaviour. Nature 2001;409:391–395. [PubMed: 11201753]

10. Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. Journal of Physical Chemistry A 2000;104:1876–1889.

11. Goss PJE, Peccoud J. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proc Natl Acad Sci U S A 1998;95:6750–6755. [PubMed: 9618484]

12. Kastner J, Solomon J, Fraser S. Modeling a Hox gene network in silico using a stochastic simulation algorithm. Developmental Biology 2002;246:122–131. [PubMed: 12027438]

13. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. Biophysical Journal 2001;81:3116–3136. [PubMed: 11720979]

14. Rao CV, Arkin AP. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. Journal of Chemical Physics 2003;118:4999–5010.

15. Simpson ML, Cox CD, Sayler GS. Frequency domain analysis of noise in autoregulated gene circuits. Proc Natl Acad Sci U S A 2003;100:4551–4556. [PubMed: 12671069]

16. Smolen P, Baxter DA, Byrne JH. Effects of macromolecular transport and stochastic fluctuations on dynamics of genetic regulatory systems. American Journal of Physiology-Cell Physiology 1999;277:C777–C790.

17. Levchenko A. Dynamical and integrative cell signaling: challenges for the new biology. Biotechnol Bioeng 2003;84:773–782. [PubMed: 14708118]

18. Haseltine EL, Rawlings JB. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. Journal of Chemical Physics 2002;117:6959–6969.

19. Haseltine EL, Rawlings JB. On the origins of approximations for stochastic chemical kinetics. Journal of Chemical Physics 2005;123:164115. [PubMed: 16268689]

20. Puchalka J, Kierzek AM. Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. Biophysical Journal 2004;86:1357–1372. [PubMed: 14990466]

21. Fricke T, Wendt D. The Markov Automaton - A new algorithm for simulating the time-evolution of large stochastic dynamic-systems. International Journal of Modern Physics C-Physics and Computers 1995;6:277–306.

22. Elf, J.; Doncic, A.; Ehrenberg, M. Mesoscopic reaction-diffusion in intracellular signaling. SPIE's "First International Symposium on Fluctuations and Noise"; 2003. p. 114-124.

23. Elf J, Ehrenberg M. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Research 2003;13:2475–2484. [PubMed: 14597656]

24. Resat H, Wiley HS, Dixon DA. Probability-weighted dynamic Monte Carlo method for reaction kinetics simulations. J Phys Chem B 2001;105:11026–11034.

25. Resat H, Ewald JA, Dixon DA, Wiley HS. An integrated model of epidermal growth factor receptor trafficking and signal transduction. Biophysical Journal 2003;85:730–743. [PubMed: 12885624]

26. Gillespie DT, Petzold LR. Improved leap-size selection for accelerated stochastic simulation. Journal of Chemical Physics 2003;119:8229–8234.

27. Rathinam M, Petzold LR, Cao Y, Gillespie DT. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. Journal of Chemical Physics 2003;119:12784–12794.

28. Cao Y, Gillespie DT, Petzold LR. Efficient step size selection for the tau-leaping simulation method. Journal of Chemical Physics 2006;124:044109. [PubMed: 16460151]

29. Chatterjee A, Mayawala K, Edwards JS, Vlachos DG. Time accelerated Monte Carlo simulations of biological networks using the binomial tau-leap method. Bioinformatics 2005;21:2136–2137. [PubMed: 15699024]

30. Chatterjee A, Vlachos DG, Katsoulakis MA. Binomial distribution based tau-leap accelerated stochastic simulation. Journal of Chemical Physics 2005;122:024112. [PubMed: 15638577]

31. Tian TH, Burrage K. Binomial leap methods for simulating stochastic chemical kinetics. Journal of Chemical Physics 2004;121:10356–10364. [PubMed: 15549913]

32. Pettigrew MF, Resat H. A multinomial tau-leaping method for stochastic kinetic simulations. J Chem Phys 2007;126:084101. [PubMed: 17343434]

33. Stundzia AB, Lumsden CJ. Stochastic simulation of coupled reaction-diffusion processes. Journal of Computational Physics 1996;127:196–207.

34. Burke P, Schooler K, Wiley HS. Regulation of epidermal growth factor receptor signaling by endocytosis and intracellular trafficking. Mol Biol Cell 2001;12:1897–1910. [PubMed: 11408594]

35. Ozcelik S, Orr G, Hu D, Chii-Shiarng C, Resat H, Harms GS, Opresko LK, Wiley HS, Colson SD. FRET measurements between small numbers of molecules identifies subtle changes in receptor interactions. Proceedings of the International Society of Optical Engineering 2004;5323:119–127.

36. Viollier PH, Thanbichler M, McGrath PT, West L, Meewan M, McAdams HH, Shapiro L. Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. Proc Natl Acad Sci USA 2004;101:9257–9262. [PubMed: 15178755]

37. McAdams HH, Shapiro L. A bacterial cell-cycle regulatory network operating in time and space. Science 2003;301:1874–1877. [PubMed: 14512618]

38. Judd EM, Ryan KR, Moerner WE, Shapiro L, McAdams HH. Fluorescence bleaching reveals asymmetric compartment formation prior to cell division in *Caulobacter*. Proc Natl Acad Sci USA 2003;100:8235–8240. [PubMed: 12824468]

39. Gardiner CW, McNeil KJ, Walls DF, Matheson IS. Correlations in stochastic theories of chemical reactions. Journal of Statistical Physics 1976;14:307–331.

40. Chaturvedi S, Gardiner CW, Matheson IS, Walls DF. Stochastic analysis of a chemical reaction with spatial and temporal structures. Journal of Statistical Physics 1977;17:469–489.

41. Gillespie, DT. Markov Processes: An Introduction for Physical Scientists. Academic Press; 1992.

42. Pettigrew MF, Resat H. Modeling signal transduction networks: a comparison of two stochastic kinetic simulation algorithms. J Chem Phys 2005;123:114707. [PubMed: 16392583]

43. Ascher, UM.; Petzold, LR. Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. SIAM; Philadelphia: 1998.

44. McQuarrie DA. Stochastic approach to chemical kinetics. J Appl Prob 1967;4:413–478.

45. Gibson, MA.; Bruck, J. California Institute of Technology Report ETR026. 1998.

46. Cao Y, Li H, Petzold LR. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. Journal of Chemical Physics 2004;121:4059–4067. [PubMed: 15332951]

47. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. Journal of Chemical Physics 2001;115:1716–1733.

48. Cao Y, Gillespie DT, Petzold LR. Avoiding negative populations in explicit Poisson tau-leaping. Journal of Chemical Physics 2005;123:054104. [PubMed: 16108628]

49. Gillespie DT. The chemical Langevin equation. Journal of Chemical Physics 2000;113:297–306.

50. Gillespie DT. The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction. Journal of Physical Chemistry A 2002;106:5063–5071.

51. Gillespie DT. The multivariate Langevin and Fokker-Planck equations. American Journal of Physics 1996;64:1246–1257.

52. Morton-Firth, CJ. PhD thesis. Cambridge, UK: University of Cambridge; 1998.

53. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. Bioinformatics 2004;20:3289–3291. [PubMed: 15217809]

54. Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J, Kitano H. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. Omics 2003;7:355–372. [PubMed: 14683609]

55. Adalsteinsson D, McMillen D, Elston TC. Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks. BMC Bioinformatics 2004;5:24. [PubMed: 15113411]

56. Funahashi A, Tanimura N, Morohashi M, Kitano H. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. Bio Silico 2003;1:159–162.

57. Kitano H. A graphical notation for biochemical networks. Bio Silico 2003;1:169–176.

58. Shapiro BE, Levchenko A, Meyerowitz EM, Wold BJ, Mjolsness ED. Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. Bioinformatics 2003;19:677–678. [PubMed: 12651737]

59. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U. COPASI--a COmplex PAthway SImulator. Bioinformatics 2006;22:3067–3074. [PubMed: 17032683]

60. Ramsey S, Orrell D, Bolouri H. Dizzy: stochastic simulation of large-scale genetic regulatory networks. J Bioinform Comput Biol 2005;3:415–436. [PubMed: 15852513]

61. Takahashi K, Ishikawa N, Sadamoto Y, Sasamoto H, Ohta S, Shiozawa A, Miyoshi F, Naito Y, Nakayama Y, Tomita M. E-Cell 2: multi-platform E-Cell simulation system. Bioinformatics 2003;19:1727–1729. [PubMed: 15593410]

62. Takahashi K, Kaizu K, Hu B, Tomita M. A multi-algorithm, multi-timescale method for cell simulation. Bioinformatics 2004;20:538–546. [PubMed: 14990450]

63. Dhar PK, Meng TC, Somani S, Ye L, Sakharkar K, Krishnan A, Ridwan AB, Wah SH, Chitre M, Hao Z. Grid cellware: the first grid-enabled tool for modelling and simulating cellular processes. Bioinformatics 2005;21:1284–1287. [PubMed: 15546936]

64. Stiles, JR.; Bartol, TM. Monte Carlo methods for simulating realistic synaptic microphysiology using MCell. In: De Schutter, E., editor. Computational Neuroscience: Realistic Modeling for Experimentalists. CRC Press; 2001. p. 87-127.

65. Hattne J, Fange D, Elf J. Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics 2005;21:2923–2924. [PubMed: 15817692]

66. Ander M, Beltrao P, Di Ventura B, Ferkinghoff-Borg J, Foglierini M, Kaplan A, Lemerle C, Tomas-Oliveira I, Serrano L. SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. Syst Biol (Stevenage) 2004;1:129–138. [PubMed: 17052123]

67. Cao, Y.; Petzold, LR. Trapezoidal tau-leaping formula for the stochastic simulation of bio-chemical systems. Proceedings of Foundations of Systems Biology in Engineering (FOSBE 2005); 2005. p. 149-152.

68. Schmidt H, Jirstrand M. Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics 2006;22:514–515. [PubMed: 16317076]

69. Slepchenko BM, Schaff JC, Macara I, Loew LM. Quantitative cell biology with the Virtual Cell. Trends Cell Biol 2003;13:570–576. [PubMed: 14573350]

70. Moraru II, Schaff JC, Slepchenko BM, Loew LM. The virtual cell: an integrated modeling environment for experimental and computational cell biology. Ann N Y Acad Sci 2002;971:595–596. [PubMed: 12438191]

71. Isaacson, SA.; Peskin, CS. Courant Institute of Mathematical Sciences Report. 2004. Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations.

72. Mattheyses T, Simmons M. Hybrid simulation of cellular behavior. Bioinformatics 2004;20:316–322. [PubMed: 14960457]

73. Cao Y, Gillespie DT, Petzold L. The slow-scale stochastic simulation algorihm. Journal of Chemical Physics 2005;122:014116.

Table 1

Reactions and stoichiometry matrix defining the example model

	Reaction	Propensity	Stoichiometry (v)				
			[L]	[R]	[RL]	[E]	[RLE]
(1)	L+R→RL	k_1[L][R]	−1	−1	+1	0	0
(2)	RL→R+L	k_2[RL]	+1	+1	−1	0	0
(3)	RL+E→RLE	k_3[RL][E]	0	0	−1	−1	+1
(4)	RLE→RL+E	k_4[RLE]	0	0	+1	+1	−1