ABRIDGING CHEMICAL REACTION NETWORKS: IT'S A SUBTLE BUSINESS

Dan T. Gillespie[(a)]*, Yang Cao[(b)], Kevin R. Sanft[(c)], and Linda R. Petzold[(c)]
[(a)]Dan T Gillespie Consulting, 30504 Cordoba Pl, Castaic, CA 91384
[(b)]Dept. of Computer Science, Virginia Tech, Blacksburg, VA 24061
[(c)]Dept. of Computer Science, University of California, Santa Barbara, CA 93106

Abstract

We examine, in a discrete-stochastic setting, the benefits and liabilities of replacing the three-reaction set $S_1 \rightleftharpoons S_2 \rightarrow S_3$ with a single S_3-producing reaction. We develop a novel criterion for deciding whether such an abridgment can be accomplished in a way that accurately replicates the production of S_3 molecules, and we derive a formula for estimating the consequent speedup in stochastic simulation. We show that in all cases in which such an abridgment can be done accurately and with a significant gain in simulation speed, a procedure called the slow-scale stochastic simulation algorithm provides a robust and theoretically transparent way of implementing the abridgment.

Keywords

model reduction, stochastic simulation efficiency, exponential waiting time, first passage time

Introduction

The strategy of "simplifying" a set of reactions by removing or modifying selected reactions and species goes back at least to the famous abridgment of the enzyme-substrate reactions by Michaelis and Menten (1913). They replaced the three reactions $E + S \rightleftharpoons ES \rightarrow E + P$ with the single reaction $S \rightarrow P$, working of course with the ordinary differential equations of conventional deterministic chemical kinetics. We want to examine this abridgment strategy in the more general stochastic context of stochastic chemical kinetics, the domain of the chemical master equation (CME) and the stochastic simulation algorithm (SSA). However, to keep the mathematics at a tractable level, we will focus our analysis on the simpler set of reactions,

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \xrightarrow{c_3} S_3\,, \qquad (1)$$

with c_1 and c_3 both assumed to be non-zero. Our goal will be to investigate the replacement of this set of three reactions with a *single* S_3-producing reaction.

All modelers appreciate the temptation to replace reactions (1) with some single reaction such as

$$S_1 \xrightarrow{c} S_3 \qquad (2)$$

in circumstances where the values of three reaction constants in (1) are not all known: Better to have just one unknown constant than three! But if, as we assume here, the modeler believes that (1) really describes what is going on, then choosing an optimal value for c in (2) will inevitably make assumptions about the values of the three rate constants in (1). Arguably, it would be better to use (1) with those assumptions made explicitly and openly, as that would not only preserve the topology of reactions (1), but also make it easier to incorporate later new information about the unknown rate constants. A more legitimate motive for replacing reactions (1) with a single S_3-producing reaction would be to speed up the stochastic simulation of the reactions. But the modeler should have some definite goal in mind when seeking such an abridgment, because abridging does have a downside:

* To whom all correspondence should be addressed

Replacing reactions (1) with a single reaction is always an approximation, the consequences of which might be difficult to predict. Furthermore, if reactions (1) are embedded in a larger network of reactions, some of which either produce or consume species S_1 or S_2, then we could have a serious problem modeling the whole system if the abridged reaction eliminates one of those species – as for example reaction (2) has eliminated S_2. In the following, we will adopt the position that the purpose of replacing reactions (1) with a single S_3-producing reaction is to speed up the simulation of S_3-production via the SSA. This may not be the only reasonable goal of such an abridgment, but it is surely an obvious one.

We will first develop a quantitative measure of the resulting speed up for various values of the rate constants in (1). Next we will use a novel criterion for assessing the accuracy of such an abridgment for various values of the rate constants. Finally, we will offer an answer to the two questions: Under what conditions can such an abridgment be done *accurately* and with a *substantial speed up* in stochastic simulation? And exactly *how* should the abridgment be implemented? For details of the following exposition, see Gillespie, et al. (2009).

The Gain in Simulation Speed

Since the SSA simulates reaction events one at a time, the time required for an SSA run is roughly proportional to the number of events simulated. Since a single S_3-producing reaction would require only one reaction event to produce each new S_3 molecule, the speed up in stochastic simulation achieved by replacing reactions (1) with a single S_3-producing reaction can be reckoned as the average number of reaction events that have to occur before reactions (1) produces a single S_3 molecule. We can estimate that number by reasoning as follows.

On each visit to state S_2, a molecule has probability $c_3/(c_2+c_3)$ of going on to state S_3. So in n visits to state S_2, a molecule will go on to state S_3 an average of $nc_3/(c_2+c_3)$ times. To get an average of *one* visit to state S_3, the molecule will thus have to visit state S_2 a total of n_1 times where $n_1=(c_2+c_3)/c_3$. If the molecule starts in state S_1, each of those n_1 visits to state S_2 requires the SSA to simulate exactly two reaction events, namely the R_1 reaction that brings it to state S_2, and either the R_2 or R_3 reaction event that takes it away. Therefore, the average number of reaction events required for the molecule to get from state S_1 to S_3 is $2n_1=2(c_2+c_3)/c_3$. If the molecule had instead started out in state S_2, it would be exactly one reaction event closer to reaching state S_3, so the average number of reaction events needed to produce one S_3 molecule would be $2n_1-1$. That, for our purposes, in not a significant difference. We conclude that the gain G in stochastic simulation efficiency achieved by replacing reactions (1) with a single S_3-producing reaction is

$$G=2\left(\frac{c_2+c_3}{c_3}\right). \tag{3}$$

The result (3) shows that $G\gg 1$ if and only if

$$c_2\gg c_3, \tag{4}$$

in which case $G\approx 2c_2/c_3$. If $c_2\approx c_3$ the speedup would be a factor of about 4, while if $c_2\ll c_3$ the speedup would be a factor of about 2. All these speedup factors will be diminished if reactions (1) are embedded in a larger set of reactions. We conclude that only when condition (4) is satisfied can the benefits of replacing reactions (1) with a single S_3-producing reaction outweigh the loss of accuracy and robustness entailed in such an abridgment.

Accuracy: The Importance of Being Exponential . . .

Stochastic chemical kinetics is based on the assumption that the dynamics of any chemical reaction are governed by a *propensity function*. In the case of the unimolecular reaction (2), that assumption is this: given x_1 S_1 molecules, the probability that some one of them will become an S_2 molecule via reaction (2) in the next infinitesimal time dt is given by the product of dt times the propensity function cx_1. This assumption can be shown to be mathematically equivalent to assuming that, in the absence of competing reactions, the time required for reaction (2) to fire is an exponential random variable with mean $1/cx_1$. More generally, for *any* single S_3-producing reaction, the time before the reaction fires, in the absence of competing reactions, must be an exponential random variable. It follows that a single S_3-producing reaction will be able to accurately replicate reactions (1) only if reactions (1) produce S_3 molecules in a like manner – i.e., only if, in reactions (1), the time to the next firing of reaction R_3 is an *exponential random variable*.

The utility of this observation lies in the fact that it is possible to calculate the exact probability density function (pdf) of the time-to-firing of reaction R_3 in (1). We will exhibit the results of that calculation in the next section. By then checking to see under what conditions that pdf is can be well approximated by the canonical exponential form $a\exp(-at)$, where a is some constant, we will be able to determine when it is possible to make an accurate single-reaction abridgment of reactions (1). And when that approximation *is* satisfactory, the constant a will stand as the propensity function of the surrogate reaction.

. . . Even in the Deterministic Limit!

It is interesting to note that the foregoing exponential criterion for accuracy in a stochastic setting applies equally to the deterministic setting. That is, again taking reaction (2) as an example, only if the waiting time for a single S_1 molecule to become an S_3 molecule via reaction (2) is the *exponential* random variable with mean $1/c$ will the conventional ordinary differential equation description of reaction (2) be valid in the large-population limit where deterministic chemical kinetics applies.

To demonstrate this perhaps surprising fact, let each of n S_1 molecules be assigned a "reaction time" in the form of a sample τ of the exponential random variable with mean $1/c$. This can be accomplished simply by taking $\tau = c^{-1}\ln(r^{-1})$, where r is a uniform random number in the unit interval. Now order those reaction times so that the smallest is τ_1, the next smallest is τ_2, etc. The consequent S_3 population trajectory will be the stair-step plot which is 0 from $t=0$ to $t=\tau_1$, 1 from $t=\tau_1$ to $t=\tau_2$, 2 from $t=\tau_2$ to $t=\tau_3$, …, and n from $t=\tau_n$ to $t=\infty$. A plot of an S_3 molecular population trajectory generated in this way for $n=300$ and $c=1$ is shown as the jagged curve in Fig. 1. The dashed curve is a plot of the solution to the deterministic rate equation for reaction (2), namely

$$\frac{dX_3(t)}{dt} = cX_1(t) = c\big[n - X_3(t)\big], \qquad (5)$$

for the initial condition $X_3(0)=0$. The close agreement between the two curves will improve if n is taken larger; indeed, if n were increased to 30000 the differences between the stochastic trajectory and the deterministic trajectory would be completely unnoticeable. This is to be expected since for sufficiently large molecular populations deterministic chemical kinetics provides an accurate approximation to stochastic chemical kinetics (for a new proof of that result, see Gillespie (2009)). But such agreement would *not* be obtained if the conversion time of an S_1 molecule were distributed in any way *other* than exponentially. For instance, if the time-to-reaction were uniformly distributed in the small interval $[0.9,1.1]$, a distribution that would give the same mean conversion time 1, then the stochastic trajectory would asymptotically approach the curve that is 0 in $[0,0.9]$, rises linearly to n in the interval $[0.9,1.1]$, and then stays at n thereafter.

The requirement that the time to the next R_3 reaction in (1) be approximately *exponentially* distributed thus applies to deterministic chemical kinetics as well as to stochastic chemical kinetics. This exponential requirement provides a robust criterion for accuracy in any replacement of reactions (1) by a single S_3-producing reaction.

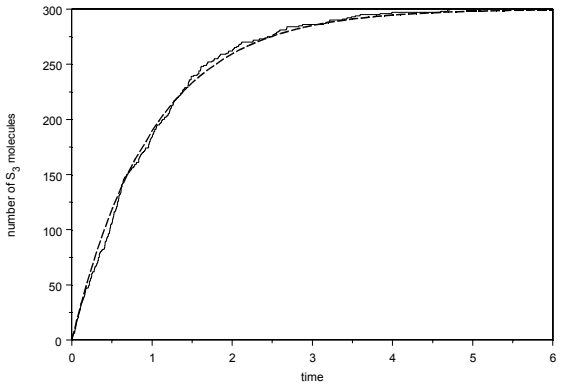

Figure 1. The jagged curve is the S_3 population trajectory obtained when each of $n=300$ S_1 molecules undergoes reaction (2) after a random waiting time that is exponentially distributed with mean $1/c=1$. The dashed curve is a plot of the corresponding solution to the deterministic reaction rate equation (5). But Eq. (5) would not provide a satisfactory fit to the behavior of this system if the time-to-conversion of each S_1 molecule were distributed other than exponentially.

The Time to an R_3 Reaction

Let $T(x_1,x_2)$ be the time to the next firing of reaction R_3 in reaction set (1) when there are x_1 S_1 molecules and x_2 S_2 molecules. An exact computation of the pdf $P(t;x_1,x_2)$ of this random variable has been carried out by Gillespie, et al. (2009). They found that

$$\begin{aligned} P(t;x_1,x_2) = {} & x_1 c_3 p(2,t|1,0)\big(p(1,t|1,0)+p(2,t|1,0)\big)^{x_1-1} \\ & \times\big(p(1,t|2,0)+p(2,t|2,0)\big)^{x_2} \\ & + x_2 c_3 p(2,t|2,0)\big(p(1,t|1,0)+p(2,t|1,0)\big)^{x_1} \\ & \times\big(p(1,t|2,0)+p(2,t|2,0)\big)^{x_2-1}. \end{aligned} \qquad (6)$$

Here, $p(n,t|\alpha,0)$ is the probability that a particular S_α molecule ($\alpha=1$ or 2) at time 0, reacting according to reactions (1), will be an S_n molecule ($n=1$, 2, or 3) at time $t>0$. It can be shown to be given by

$$p(1,t|1,0) = \frac{1}{(\lambda_+-\lambda_-)}\Big[(\lambda_+-c_1)\mathrm{e}^{-\lambda_- t} + (c_1-\lambda_-)\mathrm{e}^{-\lambda_+ t}\Big], \qquad (7a)$$

$$p(2,t|1,0) = \frac{c_1}{(\lambda_+-\lambda_-)}\Big[\mathrm{e}^{-\lambda_- t} - \mathrm{e}^{-\lambda_+ t}\Big], \qquad (7b)$$

$$p(1,t|2,0) = \frac{c_2}{(\lambda_+-\lambda_-)}\Big[\mathrm{e}^{-\lambda_- t} - \mathrm{e}^{-\lambda_+ t}\Big], \qquad (7c)$$

$$p(2,t|2,0)=\frac{1}{(\lambda_+-\lambda_-)}\left[(c_1-\lambda_-)\mathrm{e}^{-\lambda_- t}+(\lambda_+-c_1)\mathrm{e}^{-\lambda_+ t}\right], \quad (7d)$$

where

$$\lambda_\pm \equiv \tfrac{1}{2}\left[(c_1+c_2+c_3)\pm\sqrt{(c_1+c_2+c_3)^2-4c_1c_3}\right]. \quad (8)$$

We have seen that in order for reactions (1) to be replaced by a single S_3-producing reaction, it is necessary that $T(x_1,x_2)$ be approximately exponentially distributed. That formula (6) for $P(t;x_1,x_2)$ does not generally have the canonical exponential form is obvious. For example, in the simple cases of a single S_1 molecule or a single S_2 molecule, Eq. (6) gives

$$P(t;1,0)=\frac{c_1c_3}{(\lambda_+-\lambda_-)}\left[\mathrm{e}^{-\lambda_- t}-\mathrm{e}^{-\lambda_+ t}\right], \quad (9a)$$

$$P(t;0,1)=\frac{c_3}{(\lambda_+-\lambda_-)}\left[(c_1-\lambda_-)\mathrm{e}^{-\lambda_- t}+(\lambda_+-c_1)\mathrm{e}^{-\lambda_+ t}\right]. \quad (9b)$$

Plots of these two functions for $c_1=c_3=1$ and $c_2=0.1$ are shown in Fig. 2 on a semi-log scale, where the exponential form would appear as a down-sloping straight line. Neither function has that character, although $P(t;0,1)$ comes closer.

The consequences of the non-exponential character of $P(t;1,0)$ are shown in Fig. 3. Here the jagged solid line shows the S_3 population as a function of time obtained in a single SSA run of reactions (1) with $c_1=c_3=1$, $c_2=0.1$, and the initial condition $(x_1,x_2,x_3)=(300,0,0)$. The dashed line shows the average of 10000 such SSA trajectories. The dotted line shows the S_3 population computed from the deterministic reaction rate equation (5), using for c the value for which $1/c$ equals the mean conversion time for a single S_1 molecule in reactions (1) as computed from Eq. (9a). The mismatch between the dashed and dotted curves shows that a single S_3-producing reaction replacement for reactions (1) would create S_3 molecules too rapidly early on, and two slowly at later times.

Another worrisome feature of Fig. 2 is its indication that $P(t;x_1,x_2)$ depends in general on x_1 and x_2 separately, and not on only their sum. This is worrisome because a single S_3-producing reaction can generally track only x_1+x_2, so the individual values of those two variables would not be available to the simulating program.

Despite these negative indicators, Gillespie, et al. (2009) showed that there are four and only four circumstances in which $P(t;x_1,x_2)$ can be well approximated by the exponential form $a\exp(-at)$, with a depending on x_1 and x_2 only through $x_{12}\equiv x_1+x_2$. Specifically, in either of the two cases

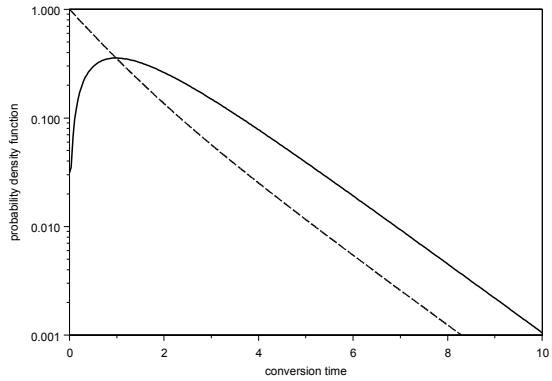

Figure 2. Plots of the pdfs of $T(1,0)$ (solid curve) and $T(0,1)$ (dashed curve) in Eqs. (9) for $c_1=c_3=1$ and $c_2=0.1$. An exponential distribution would show a down-sloping straight line. Note the change in slope of the dashed curve around $t=2$.

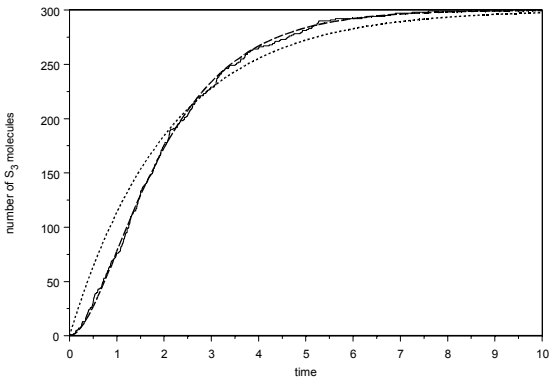

Figure 3. The jagged line is from an SSA simulation of reactions (1) for $c_1=c_3=1$, $c_2=0.1$, and the initial state $(x_1,x_2,x_3)=(300,0,0)$. The dashed line is the average of 10000 such trajectories. The dotted curve is the average trajectory of a single S_3-producing reaction that has the same average production rate. The mismatch between the dashed and dotted curves is a consequence of the non-exponential character of formula (9a).

$$c_2 \gg c_1, \quad (10a)$$

$$c_3 \gg c_1, \quad (10b)$$

$P(t;x_1,x_2)$ can be well approximated by the exponential form with decay constant

$$a = \frac{c_1 c_3 x_{12}}{c_2 + c_3}, \tag{11}$$

with the understanding that $x_2 \approx 0$ (which is an inevitable consequence of c_1 being small), and also with the understanding that we are not interested in phenomena occurring on timescales smaller than $(c_2 + c_3)^{-1}$. And in either of the two cases

$$c_1 \gg c_3, \tag{12a}$$

$$c_2 \gg c_3, \tag{12b}$$

$P(t; x_1, x_2)$ can be well approximated by the exponential form with decay constant

$$a = \frac{c_1 c_3 x_{12}}{c_1 + c_2}, \tag{13}$$

with no restrictions on x_2, but with the understanding that we are not interested in phenomena occurring on timescales smaller than $(c_1 + c_2)^{-1}$.

The result (11) has previously been derived by Mastny, et al. (2007) under the assumption that both conditions (10a) and (10b) are satisfied. And as we will discuss more fully below, the result (13) has previously been derived by Cao, et al. (2005) under essentially condition (12b). But note that the four conditions (10) and (12) are not mutually exclusive; e.g., if $c_2 \gg c_3 \gg c_1$, both conditions (10b) and (12b) are satisfied, and Eqs. (11) and (13) both reduce to $a = c_1 c_3 x_{12} / c_2$. Nor do conditions (10) and (12) cover the full parameter space of reactions (1); e.g., if $c_1 = c_2 = c_3$, none of conditions (10) and (12) are satisfied.

Abridgment With A Purpose

Several special cases of the results in the preceding section are fairly obvious. For example, if $c_2 = 0$ and condition (12a) is satisfied, Eq. (13) implies that an abridgment should be possible with $a = c_3 x_{12}$. This can be understood by observing that when $c_1 \gg c_3$, all S_1 molecules are very quickly converted to S_2 molecules, and the resulting pool of x_{12} S_2 molecules is then converted to S_3 molecules with rate constant c_3. But according to Eq. (3), making this abridgment of reactions (1) would speed up an SSA run by only a factor of 2, and even less if reactions (1) were embedded in a larger set of reactions. It seems likely that such a modest gain in simulation speed would make implementing the abridgment worthwhile, especially since the collateral assumption that the S_1 population is always 0 would pose a problem if S_1 molecules were involved in the other reactions.

Recall that our aim was to identify circumstances in which the abridgment of reactions (1) would not only be accurate, but would also produce a significant speedup in stochastic simulation. We showed earlier that the latter benefit can be realized only under condition (4). We therefore conclude that, of the four conditions (10a), (10b), (12a) and (12b) that are amenable to accurate abridgment, *only condition (12b) meets our requirement for a purposeful abridgment*.

But implementing that abridgment is not as straightforward as one might think. Thus, suppose we simply replaced reactions (1) with reaction (2) using the propensity function (13). Figure 4 shows an SSA simulation of that surroget reaction for the rate constant values and initial conditions

$$c_1 = 3,\ c_2 = 2,\ c_3 = 10^{-4}, \tag{14a}$$

$$X_1(0) = 300,\ X_2(0) = X_3(0) = 0. \tag{14b}$$

Figure 5 shows the results of an SSA simulation of reactions (1) for these same parameter values, with the populations plotted only after each R_3 reaction event.

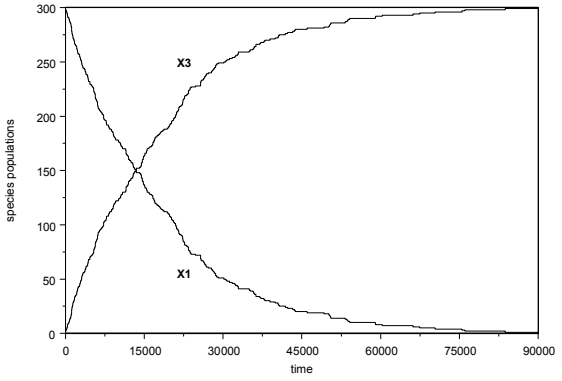

Figure 4. Showing results of an SSA run of reaction (2) with the propensity function (13) given the parameter values (14).

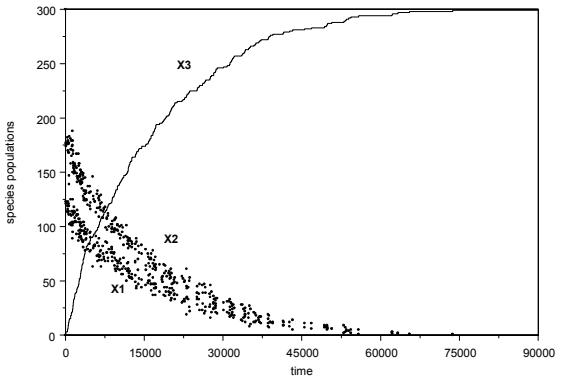

Figure 5. Showing results of an SSA run of reactions (1) using the parameter values (14), plotting the species populations immediately after each R_3 reaction.

Whereas the SSA run in Fig. 4 required simulating 300 reaction events, the SSA run in Fig. 5 required simulating 1.2×10^7 reaction events. The simulation speedup resulting from the abridgment is therefore by a very substantial factor of 4×10^4, exactly as predicted by Eq. (3). A comparison of Figs. 4 and 5 shows that the replacement reaction (2) does an excellent job of replicating the time evolution of the S_3 population in reactions (1). But the abridgment incorrectly represents the S_1 population, and it does not represent the S_2 population at all. These failings could pose a problem if reactions (1) were embedded in a larger set of reactions, some of which involved species S_1 or S_2.

A Robust Recipe for Purposeful Abridgment

The *slow-scale stochastic simulation algorithm* (ssSSA) is a procedure for simulating stochastic chemical systems that are "stiff", in that they have a wide separation of timescales and the fastest mode is stable. The ssSSA was developed by Cao, et al. (2005) as a refinement of methods introduced earlier by Haseltine and Rawlings (2002) and Rao and Arkin (2003). Cao, et al. (2005) showed that if in reactions (1) R_1 and R_2 were "fast" and R_3 was "slow" – descriptors that will be defined more precisely below – the ssSSA allows one to essentially skip over all the R_1 and R_2 events and simulate *only* the R_3 events, *provided* we use for R_3 the propensity function (13). The full ssSSA procedure for reactions (1) goes as follows:

1° In state (x_1,x_2,x_3) at time t, and with $x_{12}=x_1+x_2$, evaluate $\bar{a}_3=c_1c_3x_{12}/(c_1+c_2)$.

2° Generate the time τ to the next R_3 event as sample of the exponential random variable with mean $1/\bar{a}_3$.

3° Actualize the next R_3 event by replacing $t\leftarrow t+\tau$, $x_3\leftarrow x_3+1$, and $x_{12}\leftarrow x_{12}-1$.

4° Set x_2 equal to a sample of the binomial random variable $\mathcal{B}\left(c_1/(c_1+c_2),x_{12}\right)$, and then set $x_1=x_{12}-x_2$.

5° Record $\left(t,x_1,x_2,x_3\right)$. Return to step 1°, or else stop.

Figure 6 shows the result of executing this simulation procedure using the same parameter values (14) as in Figs. 4 and 5. In common with the simulation in Fig. 4, this ssSSA run gives an accurate rendering of the time behavior of the S_3 population, and it did so by simulating only 300 reaction events (although with a slightly longer execution time because of step 4°). But unlike the simulation in Fig. 4, the ssSSA run also gives an excellent rendering of the time behavior of the S_1 and S_2 populations, at least on the timescale of these figures.

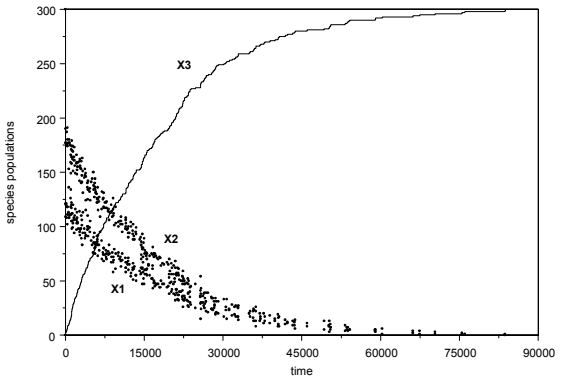

Figure 6. Showing results of an ssSSA run of reactions (1) using the same parameter values (14) as in Figs. 4 and 5.

The criterion for R_1 and R_2 to be "fast" and R_3 to be "slow" was incorrectly stated in Cao, et al. (2005). As was later shown in Gillespie, et al (2009), the only condition needed to secure those designations is $c_2\gg c_3$, i.e., condition (12b). The *physical* requirement is that, between successive firings of R_3, there should usually be $\gg1$ firings of R_1 and R_2. By the argument given in our derivation Eq. (3), condition (12b) is a sufficient to ensure that. Surprisingly, the magnitude of c_1 (which we have assumed throughout to be non-zero) plays absolutely no role in determining whether or not R_1 is a "fast" or "slow" reaction. That's because the topology of reactions (1) is such that, when $c_2\gg c_3$, R_1 will necessarily fire as often as R_2. Again, it is the *frequency of occurrence* of a reaction, not the size of its rate constant or propensity function, that determines whether the reaction is "fast" or "slow" for the purposes of the ssSSA.

The ssSSA provides a better way of implementing a single-reaction abridgment of reactions (1) under condition (12b) than does reaction (2). The advantage of the ssSSA is its ability to correctly deal with the species S_1 and S_2. A subtle point in that connection is that the values for x_1 and x_2 computed in step 4° are used only for plotting, and step 4° can be omitted without impairing the accuracy of the simulation in any way if plots of those two species are not needed. But if reactions (1) are embedded in a larger set of reactions, all of which are "slow" with respect to R_1 and R_2, the wSSA provides a recipe for constructing "effective" propensity functions for all reactions that involve R_1 or R_2 as reactants, analogous to its effective propensity function (13) for R_3. For example, the slow reaction $S_1+S_4\xrightarrow{c_4}S_5$ with propensity function $a_4=c_4x_1x_4$ would be simulated in the ssSSA with

propensity function $\bar{a}_4 = c_4 c_2 x_{12} x_4 / (c_1 + c_2)$. See Gillespie, et al. (2009) for a full discussion of this point.

Conclusions

We have seen that the problem of replacing the reaction set (1) with a single S_3-producing reaction is surprisingly subtle with respect to *motivation*, *justification*, and *implementation*.

As to motivation, since such an abridgment is inevitably an approximation, and may have the unintended effect of complicating the simulation of reactions (1) along with other reactions that have species S_1 and S_2 as reactants, then one should have a clear goal in mind for the abridgment. The goal we adopted here, which we think will be a common one, is to effect a significant speed up in stochastic simulation. We showed that the only circumstance in which that goal can be achieved for reactions (1) is when $c_2 \gg c_3$.

Assuming that reactions (1) truly describe how S_3 molecules are produced, then replacing them with a single S_3-producing reaction will be justified only if relevant features of how reactions (1) produce S_3 molecules are approximately preserved. We pointed out that since any single S_3-producing reaction will necessarily have an exponentially distributed waiting time to fire, then in order for such a reaction to accurately replace reactions (1), the waiting time for R_3 in reactions (1) must likewise be exponentially distributed, at least approximately. And we showed that this condition applies even in the deterministic setting of large molecular populations. By carrying out a detailed "first-passage-time" analysis of the waiting time distribution for reaction R_3, we concluded that the exponential distribution requirement will be satisfied if and only if at least one of the following four conditions holds: $c_2 \gg c_1$, $c_3 \gg c_1$, $c_1 \gg c_3$, $c_2 \gg c_3$. We therefore concluded that the only circumstance in which an abridgment can be made accurately *and* with a significant speedup in stochastic simulation is when $c_2 \gg c_3$.

We next showed that implementing the abridgment when $c_2 \gg c_3$ by simply swapping reactions (1) for some single S_3-producing reaction, like (2) for example, does not generally take adequate account of species S_1 and S_2: A simple swap cannot accurately describe the behavior of those two species on the timescale of the "slow" reaction R_3; moreover, when reactions (1) are embedded in a larger set of reactions, such a replacement fails to provide a clear procedure for simulating other reactions that involve S_1 and S_2 as reactants.

We showed that an effective remedy for these deficiencies is provided by the "slow-scale stochastic simulation algorithm" (ssSSA). In an ssSSA simulation, the "fast" reactions R_1 and R_2 are skipped over, and only the "slow" reaction R_3 is simulated, along with any other slow reactions that might also be occurring. But all these slow reactions are simulated with modified propensity functions, the forms of which are prescribed by the theory underlying the ssSSA. The consistency of this ssSSA procedure has been verified in this paper by our finding that the modified propensity function the ssSSA prescribes for reaction R_3 is exactly the same as the propensity function predicted by the first-passage time analysis of the R_3 waiting time distribution for the case $c_2 \gg c_3$.

Details of all the arguments presented here may be found in Gillespie, et al. (2009). We believe that our findings for the simple reaction set (1) carry lessons for more complicated reaction sets, such as the classical enzyme-substrate reaction set, and we plan to explore those lessons in future studies.

Acknowledgments

DG was supported by the California Institute of Technology through Consulting Agreement 102-1080890 pursuant to Grant R01GM078992 from the National Institute of General Medical Sciences, and through Contract 82-1083250 pursuant to Grant R01EB007511 from the National Institute of Biomedical Imaging and Bioengineering, and also from the University of California at Santa Barbara under Consulting Agreement 054281A20 pursuant to funding from the National Institutes of Health. YC was supported by the National Science Foundation under award CCF-0726763, and also the National Institutes of Health under awards GM073744 and GM078989. KS and LP were supported by Grant R01EB007511 from the National Institute of Biomedical Imaging and Bioengineering, Pfizer Inc., DOE DE-FG02-04ER25621, NSF IGERT DG02-21715, and the Institute for Collaborative Biotechnologies through Grant DFR3A-8-447850-23002 from the U.S. Army Research Office. KS was also supported by a National Science Foundation Graduate Research Fellowship.

References

Cao, Y., Gillespie, D., Petzold, L. (2005). *J. Chem. Phys. 122*, 014116.

Gillespie, D. (2007). *Ann. Rev. Phys. Chem. 58*, 35.

Gillespie, D., Cao, Y., Sanft, K, Petzold, L. (2009). *J. Chem. Phys. 130*, 064103.

Haseltine, E., Rawlings, J. (2002). *J. Chem. Phys. 117*, 6959.

Mastny, E., Haseltine, E., Rawlings, J. (2007). *J. Chem. Phys. 127*, 094106.

Michaelis, L., Menten, M. (1913). *Biochem. Z. 49*,333.

Rao, C., Arkin, A. (2003). *J. Chem. Phys. 118*, 4999.