Probability-Weighted Dynamic Monte Carlo Method for Reaction Kinetics Simulations

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · NOVEMBER 2000

Impact Factor: 3.3 · DOI: 10.1021/jp011404w

3 AUTHORS:

H. Resat
Washington State University
47 PUBLICATIONS **1,235** CITATIONS
SEE PROFILE

Henry Steven Wiley
Pacific Northwest National Laboratory
181 PUBLICATIONS **8,920** CITATIONS
SEE PROFILE

David A Dixon
University of Alabama
768 PUBLICATIONS **22,231** CITATIONS
SEE PROFILE

Probability-Weighted Dynamic Monte Carlo Method for Reaction Kinetics Simulations

Haluk Resat,* H. Steven Wiley, and David A. Dixon

Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352

Received: April 16, 2001; In Final Form: July 9, 2001

The reaction kinetics underlying the dynamic features of physical systems can be investigated by using various approaches such as the Dynamic Monte Carlo (DMC) method. The usefulness of the DMC method to study reaction kinetics has been limited to systems where the underlying reactions occur with similar frequencies, i.e., similar rate constants. However, many interesting physical phenomena involve sets of reactions with a wide range of rate constants leading to a broad range of relevant time scales. Widely varying reaction rates result in a highly skewed reaction occurrence probability distribution. When the reaction occurrence probability distribution has a wide spectrum, the reactions with faster rates dominate the computations, making the reliable statistical sampling cumbersome. We have developed a probability-weighted DMC method by incorporating the weighted sampling algorithm of equilibrium molecular simulations. This new algorithm samples the slow reactions very efficiently and makes it possible to simulate in a computationally efficient manner the reaction kinetics of physical systems in which the rates of reactions vary by several orders of magnitude. We validate the probability-weighted DMC algorithm by applying it to two model systems: a simple chemical reaction system and a model of vesicular trafficking in living cells.

Introduction

Every physical reaction system has an inherently dynamical character independent of it being in an equilibrium or nonequilibrium state. A variety of simulation approaches can be used to treat dynamical systems. Monte Carlo (MC) simulations are widely used to compute the properties of physical systems in condensed phases and in the gas phase.[1] Although MC methods are usually employed to obtain the equilibrium properties of systems, they can also be used to study dynamical systems, and their use has been extended to the study of time-dependent properties.[2−5] In the traditional computational approach to chemical kinetics, one starts with a set of coupled ordinary differential equations (reaction rate equations) that describe the time-dependent concentration of chemical species. One then uses some integrator to calculate the concentrations as a function of time given the rate constants and a set of initial concentrations. Gillespie has shown that this formal deterministic approach can be translated to a stochastic scheme giving rise to the stochastic Dynamic Monte Carlo (DMC) approach and has shown in detail how one can connect the traditional chemical kinetics and stochastic approaches.[3,4] Assuming that the system is well mixed, the rate constants appearing in these two methods are related.[3] Because the molecules forming the physical system are complete chemical entities, they must participate in the reactions as integer species. The traditional approach based on the continuum treatment of chemical kinetics can lead to misrepresentation of the physical system, particularly when the number of molecules that can take part in the reactions is small, thus making the discrete nature of the system important. As argued by Gillespie, the use of a discrete representation where molecules are formed or consumed in integer units is more appropriate in kinetic simulation studies.[3] The sequence of events that occur in a DMC simulation can be interpreted as a series of Poisson processes.[3,5] Using the connection between DMC simulations and Poisson statistics, Fichthorn and Weinberg have shown that a physical meaning can be attached to the propagated time in DMC simulations.[5] The time elapsed in each move of a DMC simulation is not uniform,[2,3,5] and the assumption of the uniform time step can lead to inconsistent results.[6]

An important factor for consideration in any dynamical simulation is the time scale or set of time scales of the studied processes. This time scale, as well as the available computational resources, determines whether a particular simulation method can be used. For example, in the Molecular Dynamics (MD) method, the time step of integrating the equations of motion is chosen according to the fastest motions. Thus, if the motions of the bonds involving hydrogen atoms are explicitly included, the time steps on the order of 1 fs have to be used. This typically limits the total length of the simulations to 1 to 100 ns for a protein with the current computational technology. Similarly, in the DMC method, the probabilities of the fast reactions (i.e., reactions with large rate constants) are generally larger than the comparatively slower reactions. Since they carry larger probabilities, high-frequency reactions occur much more often and dominate the DMC simulations.[7−9] In general, the accuracy of a molecular simulation is determined by the reliability of its statistical sampling. For example, if certain reactions occur only a few times in a stochastic kinetics simulation, the resulting statistical accuracy would be very low, making the predictions of the simulation study unreliable. For this reason, if the involved reactions occur on broadly different time scales, the DMC simulations can run into statistical accuracy problems due to the unreliability of the sampling. This has severely limited the use of the method.

In most instances, a complete understanding of the set of reactions requires the explicit consideration of the slow and the fast reactions. For example, cells respond to their environment

* Corresponding author: Haluk Resat, Pacific Northwest National Laboratory, P.O. Box 999, mail stop: K1-83, Richland WA 99352. E-mail: haluk.resat@pnl.gov

10.1021/jp011404w CCC: $20.00
Published on Web 10/16/2001

by the interaction of ligand molecules with appropriate cell membrane receptors.[10] This requires the diffusion and binding of the ligand to the receptor. Ligand binding induces a physical change, such as phosphorylation, in the receptor. This in turn causes the receptor to interact with various components in the cell cytoplasm to transfer information to downstream elements. The entire signaling process consists of steps that range from very fast processes on molecular time scales to processes that occur on the order of minutes. Other well-studied examples are the surface adsorption/desorption of molecules during thin film growth[7] and surface catalysis.[8,9] The time scale of the reaction cascade that follows surface adsorption can be quite different from the time scale of the adsorption process. In addition, different molecules attach to a surface with widely different affinities, again giving rise to the time scale problem and computational difficulties when studying the growth process by molecular simulation methods.[7] Such difficulties are particularly obvious when the reaction following the surface adsorption involves diffusion of molecules that occurs on a very fast time scale.[8,9]

There have been a number of attempts to improve the efficiency of the DMC simulations,[2,6−9,11−13] but they have met with only limited success. We introduce a probability-weighted Dynamic Monte Carlo (PW-DMC) method to simulate kinetic reaction schemes with widely varying reaction rates. The PW-DMC method is based on the weighted sampling of the reactions as the Markov chain is generated. This requires correcting the time propagation with the weight of the preferential sampling. The weighted sampling method[1,14] has been used in equilibrium Monte Carlo simulations for over three decades but has not previously been utilized in DMC simulations. The new algorithm is applied to two model reaction systems where the reaction rates cover a time span of four orders of magnitude. The results of the new algorithm agree very well with the ordinary implementation of the DMC method, and the timing comparison shows that the PW-DMC method is computationally much more efficient than the ordinary DMC algorithm.

Methods

Dynamic Monte Carlo. In the usual implementation of DMC for kinetic simulations, each reaction is considered as an event and each event has an associated probability of occurring. The probability that a certain chemical reaction takes place in a given time interval is proportional to an effective rate constant and to the number of chemical species that can take part in that event.[3,4] For example, the probability of the first-order reaction $X \rightarrow Y + Z$ would be $a_1 N_x$ where N_x is the number of species X, and a_1 is the rate constant of the reaction. Similarly, the probability of the reverse second-order reaction $Y + Z \rightarrow X$ would be $a_2 N_Y N_Z$. As the method is a probabilistic approach based on "events", "reactions" included in the DMC simulations do not have to be solely chemical reactions. Any process that can be associated with a probability can be included as an event in the DMC simulations. For example, a substrate attaching to a solid surface can initiate a series of chemical reactions. When building the model, one can assume that the substrate is already attached to the surface and start the simulations of the subsequent chemical reaction kinetics. However, since the impact of the substrate can have an effect, a more complete modeling effort would also incorporate the initial stage of substrate arrival to the surface and its attachment. In this case, the substrate arrival (which can be due to plasma spray, free diffusion in solution, efflux from another surface, etc.) is a physical event, not a chemical event. However, since a probability can be assigned to the physical event, the arrival of the substrate can easily be included in the DMC simulations.

Another good example is taken from the biological cell signaling network that we have been studying.[15] To briefly summarize this network, epidermal growth factor receptors on the cell surface are internalized through the process of endocytosis. During endocytosis, the cell membrane invaginates to form a vesicle. The receptor-containing vesicles are transported to the interior of the cell.[16] Both the cell membrane and internalized receptors can be active, i.e., they can transfer a signal to downstream elements (mainly protein kinase substrates) through a set of biochemical reactions.[16,17] Previous work has shown that receptor internalization and endocytic vesicle trafficking are essential aspects of the signaling process and must be included in any accurate model of cell receptor signaling networks.[16] In this example, even though it is not a simple chemical reaction, the formation of an endocytic vesicle can be included in the DMC simulation studies by considering it as an event and assigning an appropriate probability. The ability to include nonchemical reactions in kinetic simulations is a very important advantage of the DMC methods over analogous approaches. Unfortunately, because of the high computational expense of involving events with widely different time scales, studies of such multistage models have been avoided in the past. The computational speed increase achieved by using our new algorithm makes it possible to investigate the reaction kinetics of such systems despite the widely different time scales of their component processes.

The basic outline of a DMC method, the direct method of Gillespie,[3] can be summarized as follows. (i) Generate a list of the components/species and define the initial distribution at time $t = 0$. (ii) Generate a list of the possible events. As discussed above, this list can include chemical reactions (either fundamental or lumped), as well as all other physical processes that can be included in the kinetic model. In making the list, a reversible chemical reaction must be separated into forward and reverse reactions, and these two reactions are treated as unrelated events. (iii) Using the current component/species distribution, prepare a probability table of all the events that can take place. Compute the total probability P_{tot} by summing the probabilities of all individual events $P(E_i)$ where $P(E_i)$ is the probability of the ith event. (iv) Pick two random numbers r_1 and r_2 from a uniform distribution in the unit interval to decide which event E_μ occurs next and the amount of time elapsed since the most recent reaction. Using one of the random numbers and the probability table, the event E_μ is determined by finding the event that satisfies the relation

$$\sum_{i=1}^{\mu-1} P(E_i) < r_1 P_{tot} \leq \sum_{i=1}^{\mu} P(E_i) \qquad (1)$$

The second random number is used to obtain the amount of time τ between the reactions

$$\tau = -(1/P_{tot}) \ln(r_2) \qquad (2)$$

As the total probability of the events changes in time, the time step between occurring events varies. Steps (iii) and (iv) are repeated at each step of the simulation. Noting that DMC is stochastic in character, the simulation runs need to be repeated many times and the results of the runs are averaged to obtain reliable data. The necessary number of runs depends on the inherent noise of the system and on the desired statistical accuracy.

Weighted Sampling. Because of the inefficiency of the usual implementation of the Monte Carlo algorithm in sampling the low probability parts of phase space, Torrie and Valleau devised an extension of Monte Carlo simulations based on nonphysical or weighted-sampling distributions.[14] In the commonly used MC algorithm, the Markov chain is generated using the transition probabilities $\pi(i \rightarrow j)$ that are based on the physical probability distribution

$$\pi(i \rightarrow j) = \frac{p_{(i \rightarrow j)}}{\sum_k P_{(i \rightarrow k)}} \tag{3}$$

The ensemble average of a physical quantity θ is obtained by taking the arithmetic average of all the simulation runs. This choice disfavors the transitions with low probabilities. If the system characteristics depend on the events that happen less frequently, then the common implementation of MC requires extremely lengthy simulations to acquire enough statistical sampling. This statistical sampling problem can be avoided if the probability distribution is multiplied with a weight function that adjusts the sampling probability distribution such that the low probability parts of the sampling space are visited more often.[14] In the case of the weighted sampling, the Markov chain is generated by using the modified probability distribution function $P_w(i \rightarrow j) = P(i \rightarrow j)\ \mathrm{Y}(i \rightarrow j)$, where Y is the biasing weight function. Since the probability of the transition $i \rightarrow j$ is weighted with $\mathrm{Y}(i \rightarrow j)$, calculation of the ensemble average of a physical quantity θ is obtained by computing the average of θ/Y. Division of θ by Y effectively corrects for the bias introduced in the sampling probability distribution.[14]

Probability-Weighted Dynamic Monte Carlo Method. The probability-weighted Dynamic Monte Carlo (PW-DMC) method incorporates weighted sampling into the Dynamic Monte Carlo method. In this approach, the third and the fourth steps of the DMC algorithm described above are replaced with the following. (iii) Using the current component/species distribution, prepare a probability table of all the events that can take place. (iv) Define the weight factor scale and compute the inverse probability weight table $w(E_\mu) = \mathrm{Y}^{-1}(E_\mu)$ for all events. Note that the stochastic simulations mentioned here use discrete numbers of molecules, i.e., the species are produced or consumed as whole integer units. Therefore, the weight table for $w(E_\mu)$ must contain only integer values. This is achieved by partitioning the continuous probability distribution into integer valued histogram bins according to the chosen weight scale. (v) Prepare the weighted probability table $P_w(E_i) = P(E_i)/w(E_i)$. (vi) Compute the total probability by summing the weighted probabilities of all individual events $P_{\mathrm{tot}} = \Sigma P_w(E_i)$. (vii) Pick two random numbers r_1 and r_2 from a uniform distribution in the unit interval. Using the weighted probabilities $P_w(E)$ in eq 1, determine which event E_μ occurs next. Note that if $w(E_\mu)$ is not unity, because of the weighted sampling, the simulation trajectory has to be corrected by assuming that the event E_μ occurred $w(E_\mu)$ consecutive times. Although any weight factor can be used in principle, the requirement that the events occur a discrete number of times is the reason for the use of integer-valued weight factors $w(E_i)$. (viii) Propagate the time according to eq 2. Since the total probability is calculated by using the weighted probabilities (see step vi), eq 2 does not need to be corrected further.

The speed-up achieved by the PW-DMC algorithm stems from the fact that the reactions with large probabilities are allowed to occur in "bundles". As step (vii) of the algorithm states, when the inverse weight of an event is larger than one, correcting for the weighted sampling requires that the event is recorded as having occurred $w(E)$ times. This allows for larger time elapses between the Monte Carlo moves and increases the computational efficiency.

Characteristic properties of a physical system are encoded in the probability distribution function of the events that take place. Even though there is an associated intrinsic uncertainty with any given probability distribution, every properly defined probability distribution function defines a unique system with certain expected average characteristics. Any alteration of a probability distribution function in effect transforms the physical system to a different one. Bundling of the reactions in the PW-DMC, in principle, can change the order in which the events take place in a Markov chain. Since the event probabilities in general depend on the number of molecules, allowing the events to occur in bundles rather than one at a time may make the probability distribution defining the stochastic simulation inexact for short periods of time. Even though such concerns are legitimate and can be important, certain approximations, such as the PW-DMC method described here, affect the system characteristics minimally and can be employed without any significant impact on the results. As shown below, with a nonsignificant increase in the statistical error, the PW-DMC algorithm preserves the correct ensemble averages and it is considerably faster than the usual implementation of DMC.

As has been argued,[8] DMC is essentially a method to solve the master equation that rules how the probabilities of the configurations are related to each other

$$\frac{\mathrm{d}P_\alpha}{\mathrm{d}t} = \sum_\beta [W_{\alpha\beta}P_\beta - W_{\beta\alpha}P_\alpha] \tag{4}$$

where $W_{\alpha\beta}$ and P_α are the transition probability of going from configuration β to α and the probability of the configuration α, respectively. Using the master equation, the statistical average $\langle X \rangle$ of the rate of change of the property X can be expressed as[8]

$$\frac{\mathrm{d}\langle X \rangle}{\mathrm{d}t} = \sum_{\alpha\beta} W_{\alpha\beta}P_\beta [X_\alpha - X_\beta] \tag{5}$$

In the PW-DMC, one again uses eq 5 but it is rearranged as

$$\frac{\mathrm{d}\langle X \rangle}{\mathrm{d}t} = \sum_{\alpha\beta} \frac{W_{\alpha\beta}P_\beta}{w} \{w[X_\alpha - X_\beta]\} \tag{6}$$

where w is the weight factor used in the PW-DMC algorithm. As eq 6 shows, PW-DMC leaves the ensemble averages unchanged; however, the fluctuations (i.e., the factor $w[X_\alpha - X_\beta]$ in the equation) in the computed properties increase with a factor of w.

The weighted sampling used in the PW-DMC method might skew the Markov chain generation for short periods of time. Because of the noise inherent in probabilistic approaches, the stochastic simulations need to be repeated many times and averaged. Therefore, in practical implementation of the kinetic simulations, DMC runs are performed many times and the trajectories are averaged to find the expected values of the studied physical quantities. In the PW-DMC method, reactions may occur in bundles, but the order of the grouped reactions is still chosen randomly and the involved errors still follow the rules of Markov processes. Because of the fundamental property of the Markov processes that the system does not have a memory

and each Markov step is an independent statistical process, the associated statistical errors cancel out to a good degree when the simulation trajectories are averaged. In other words, the repetition of the simulation runs helps to eliminate the possible effects of the weighed sampling on the statistical fluctuations. Therefore, the expected increase in the statistical fluctuations in the PW-DMC method would not be of a major concern in most instances.

As in any statistical approach, one has to ensure that the ratio of the statistical uncertainty of a property X to the value of X itself is reasonably small for the results to be reliable. If the probability weight factor $w(E)$ is proportional to the probability of the event E, $w(E)$ would be large if $P(E)$ is large (step v in the new algorithm). Since the chemical reaction probabilities are proportional to the number of molecules and the rate constant, the chemical reactions (events) with large probabilities are either extremely fast reactions or they involve species that are abundant. Therefore, if the reason a type of chemical reaction occurs often is the abundance of participating molecular species, the approximation to the true probability distribution of events would be of the order of $w(E)/N(E)$ where $N(E)$ is the number of molecules of the species taking part in the reaction.[18] If $N(E)$ is large, this ratio would be very small, and therefore, the deviation from the true probability distribution due to weighted sampling would be negligible. If the probability of an event is large due to a very large rate constant, that event will occur much more often than the other reactions, regardless of whether weighted sampling is applied or not. Bundling these reactions would not alter the dynamics of the system. For these reasons, the effects of weighted sampling of the chemical reactions in the generation of the Markov chain are not significant.

In terms of nonchemical events, the weight of the biased sampling can be chosen according to the importance of the impact the approximation might have. We note again that, in the weighted sampling, the event weights can be chosen in any desired way without violating the statistical mechanical foundations. Therefore, if allowing a certain event to take place more than once at a step would severely distort the physical outcome, the weight factor of that event can be set to unity and would not be allowed to change. This would force that event to occur as in the normal sampling algorithm. This flexibility of being able to assign the weight factors in various ways provides the PW-DMC method with an additional advantage by making it very flexible. If a fast event E_j is not essential, it can be assigned a large $w(E_j)$ factor. Even though the event E_j would take place the same total number of times (in the probabilistic sense) during the simulation run, it would occur in large bundles and the bundled reactions would take place fewer times. Since the computational expense of the DMC simulation is proportional to how many times the code has to decide which reaction(s) would take place, bundling the reactions allows the algorithm to avoid repeated fast events, which results in a significant reduction in the computational requirement.

In our implementation of the PW-DMC, the simulation code compares the possible change in the number of molecules of a certain species δN (this value is mainly the weight factor of the reaction for A $\rightarrow$ B + C or A + B $\rightarrow$ C type reactions) to the number of particles of the species N. If $\delta N/N$ is larger than a preset value (set to 1/3 in our simulations), the code adjusts the weight factor of the reaction to make $\delta N/N$ smaller than the preset value. This ensures that the statistical fluctuations are kept within reasonable limits. In the simulations reported below, such adjustments were needed only a very few number of times. As the results for the model systems show, as expected, the bias introduced by the weighted sampling has almost no effect on the averaged DMC results.

Since both are based on grouping the events, the PW-DMC method might be considered to be similar to the n-fold method.[2] Even though the simple concept behind these two algorithms is the grouping of the reactions, the PW-DMC method is fundamentally different from the n-fold method. In the n-fold method, which was originally developed to study the Ising model of spin systems,[2,11] the standard algorithm is reorganized to group similar reactions together to increase their total probability (i.e., the hit probability of being chosen). This reorganization, and the elimination of the MC attempts that would be rejected, makes the simulation more efficient and reduces the computational time. The speed gain achieved in the n-fold algorithm is only moderate, and the algorithm requires that there are similar reactions that can be grouped. The PW-DMC method, however, reduces the computational expense by intelligently modifying the probability distribution function to sample rarely occurring events more frequently and then back-correcting the results. This results in significant speed gains, and may actually increase the reliability of the results for certain systems by sampling the rare events many more times than the traditional implementation of the DMC method could.

Results

We have tested the PW-DMC method on two model systems and compared the results with the results of the DMC method implemented in the standard way.

Model System 1. In the first model, we investigate a simple set of chemical reactions for which analytical results are available. It consists of two reversible reactions A + B $\leftrightarrow$ C and A + C $\leftrightarrow$ D. Let us assume that k_1 and k_{-1} and k_2 and k_{-2}, respectively, are the forward and reverse reaction rate constants of the first and the second reactions and define the following parameters

$$r_{12} = k_1/k_{-1} \text{ and } r = k_1k_2/k_{-1}k_{-2} \tag{7a}$$

$$A_0 = A(0) + C(0) + 2D(0) \text{ and } B_0 = B(0) + C(0) + D(0) \tag{7b}$$

where $X(0)$ is the concentration of X at time $t = 0$. By using the chemical conservation rules, the steady-state values of the species A can be found by solving the following third-order polynomial equation

$$rA^3 + (r_{12} - rA_0 + 2rB_0)\,A^2 + (1 - r_{12}A_0 + r_{12}B_0)\,A - A_0 = 0 \tag{7c}$$

The steady-state concentrations of the other species are

$$B = B_0/(1 + r_{12}A + rA^2)$$

$$C = r_{12}AB$$

$$D = rA^2B \tag{7d}$$

The rate constants $k_1 = 10^{-5}$ s^{-1}, $k_{-1} = 1.0$ molecule^{-1} s^{-2}, $k_2 = 10^{-3}$ s^{-1}, and $k_{-2} = 0.1$ molecule^{-1} s^{-2} were used in the simulations, and the initial distributions of the molecules were $A(t{=}0) = 10000$, B(0) = 2500, and C(0) = D(0) = 0. With these values, the steady-state concentrations of the species would be A = 6082.3, B = 525.2, C = 31.9, and D = 1942.9 molecules.

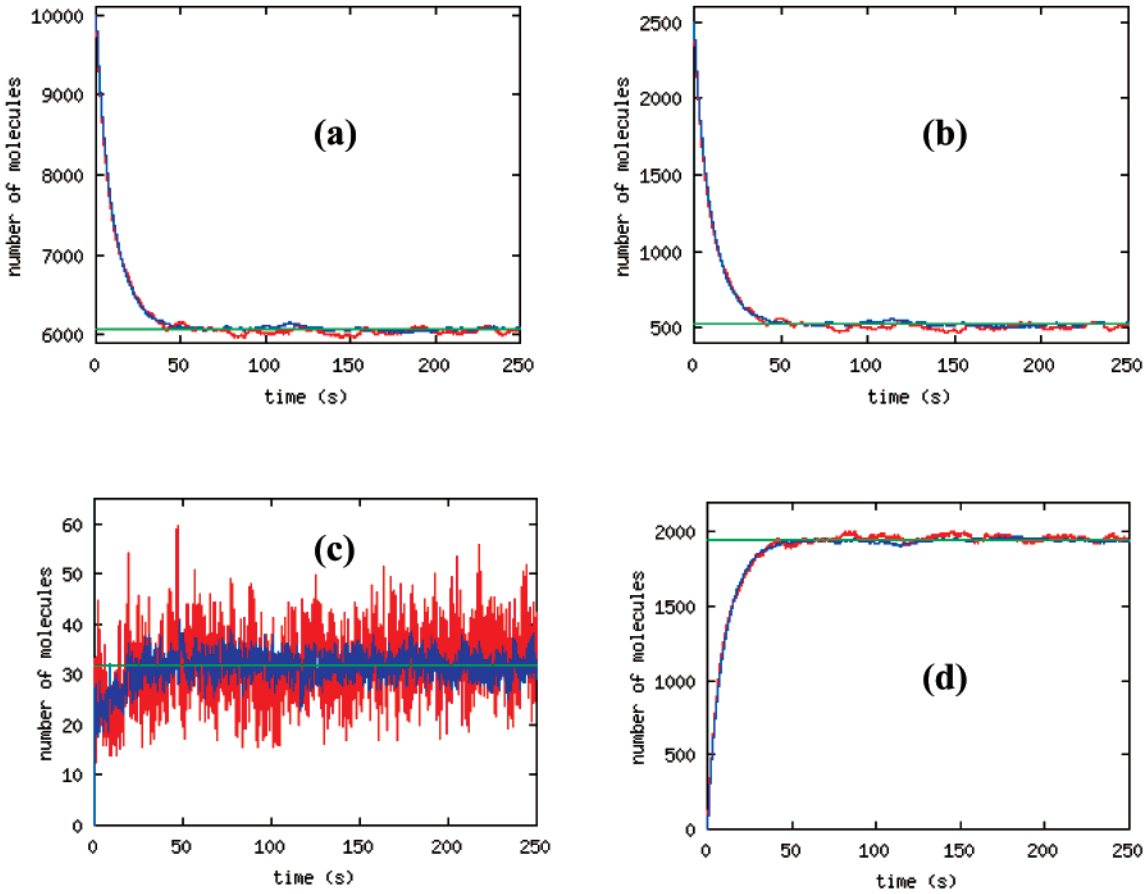

Figure 1. Number distribution of various species in the first model system. The DMC and the PW-DMC results are shown in blue and red, respectively. The green line shows the steady-state concentration of the species. The horizontal axis shows the simulation time in seconds. Graphs are for species (a) A, (b) B, (c) C, and (d) D.

Two sets of simulations were performed by using the PW-DMC and the DMC methods. Each set of simulations consisted of 12 simulation runs starting with the same initial configuration but using a different seed value for the random number generator. Figures 1(a−d) show the concentration of the species as a function of time that were computed by averaging the 12 runs for each method.[19] A bin size of 0.05 was chosen to assign the sampling weights of the studied events. In other words, events with probabilities less than 0.05 had a weight of 1, and the events with probabilities between 0.25 and 0.30 had a weight of 6, etc. The 250-second simulation length runs using the PW-DMC method took roughly 6.6 times less time to complete than the runs that used the DMC method. As the plots reveal, the results of the PW-DMC method agree very well with the results of the DMC method, and the steady state values are obtained accurately by both methods.

The model system involves two similar equations where the species A reacts with either species B or C. However, because of the differences in the rate constants, the concentrations of species B and C are very different and there are only a very small number of C molecules in the system. The agreement between the PW-DMC and the DMC methods shows by example that even under conditions of extreme number of molecule distributions, the approximations involved in the PW-DMC method do not have any serious impact on the results.

Model System 2. The second model system is a simplified version of the trafficking and signaling network in cells that we have been studying.[15] It is, however, much more complex than the simple model system investigated above. In the model system shown in Figure 2, there are A and B molecules that

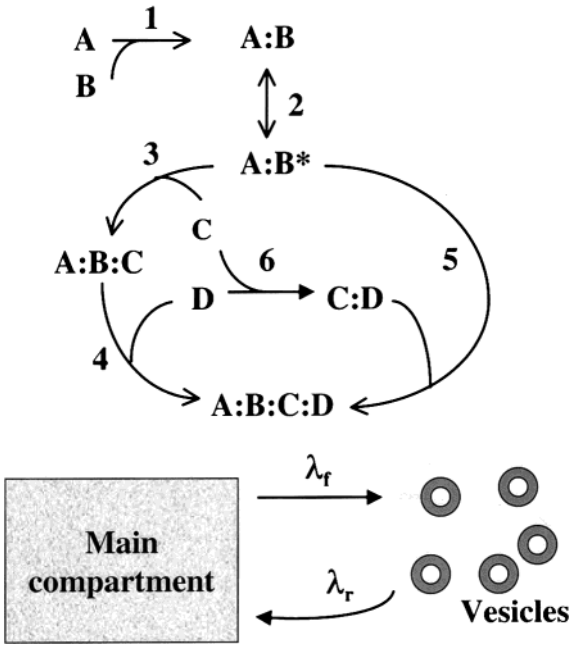

Figure 2. Reaction diagram of the second model system: the chemical reactions and the vesicle formation.

can form an A:B complex (reaction 1). In this case, for example, B is the ligand for the receptor A. The A:B complex goes through a transformation (for example, the phosphorylation of a receptor, or the conformational change which would activate the complex) and the activated A:B* species is formed (reaction

TABLE 1: List of Reactions and Number of Occurrence of the Reactions

reaction[a]		order	rate	comp	NOO[b] PW-DMC	NOO[b] DMC	% diff NOO[c]
1	A + B → A:B	2	2.8E−7	M	6619	6745	0.9
			5.6E−7	V	1513	1598	2.7
	A:B → A + B	1	4.0E−2	M	4338	4479	1.6
			2.0E−2	V	2046	2114	1.7
2	A:B → A:B*	1	0.1	M	11113	11219	0.5
			0.2	V	20351	21107	1.8
	A:B* → A:B	1	1.0E−3	M	8825	8948	0.7
			5.0E−4	V	20896	21632	1.7
3	A:B* + C → A:B:C	2	2.8E−6	M	970495	989331	1.0
	A:B:C → A:B* + C	1	5.0E−3	M	962189	981046	1.0
			5.0E−3	V	7702	7645	0.4
4	A:B:C + D → A:B:C:D	2	1.1E−6	M	1648829	1669727	0.6
	A:B:C:D → A:B:C + D	1	1.0E−3	M	1644827	1665753	0.6
			1.0E−3	V	3292	3256	0.6
5	A:B* + C:D → A:B:C:D	2	2.8E−7	M	33016	33589	0.9
	A:B:C:D → A:B* + C:D	1	2.0E−5	M	32739	33327	0.9
			1.0E−5	V	36	34	2.7
6	C + D → C:D	2	2.2E−7	M	3394945	3439664	0.7
	C:D → C + D	1	5.0E−3	M	3395430	3440180	0.7
7	vesicle formation	0	0.005		252	251	0.2
8	vesicle recycling	1	2.0E−4		227	227	0.0

[a] Order and rate are the order and the rate constant of the listed reaction. The unit of the reaction rate constant of an *n*th order reaction is molecule^{1-n}/s^n. Compartments M and V stand for the main compartment and the vesicles, respectively. Note that, since C, D, and C:D can only exist in the main compartment (see the text), certain reactions cannot take place in the vesicles. [b] NOO: Number of occurrence = the total number of times a reaction took place during the probability weighted dynamic Monte Carlo (PW-DMC) or the ordinary dynamic Monte Carlo (DMC) simulation runs. The averages of the nine runs are given. [c] Percent difference between NOOs of the PW-DMC and the DMC simulations.

TABLE 2: List of Reactant Species and Concentration Information

reactant	N[a]	VIP[b]
A	37800	1.0
B	180000	1.0
A:B		5.0
A:B*		5.0
A:B:C		0.5
A:B:C:D		0.05
C	90000	
D	54000	
C:D		

[a] Total number of molecules in the simulation. [b] Vesicle inclusion percentages of the reactant species. When a vesicle is formed, VIP percent of the reactant species that are in the main compartment at the time constitute the contents of the formed vesicle.

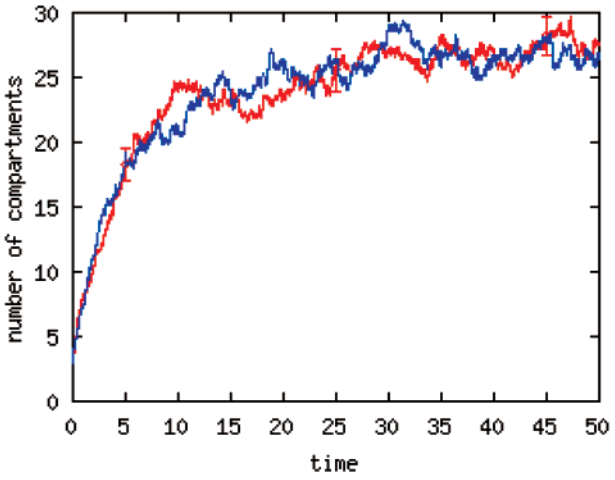

Figure 3. Total number of compartments as a function of time: DMC (blue) and PW-DMC (red). Typical error bars are shown for the PW-DMC results. The error bars are similar for the DMC.

2). The activated complex can interact with C or D molecules to form the A:B:C or the A:B:C:D complex (reactions 3, 4, and 5). In this case, C and D are, for example, the catalysts in solution or the scaffold proteins to which activated A:B* needs to bind to participate in further downstream reactions. The reaction rate constants used in this study are given in Table 1. The model test system also allows for compartment (vesicle) formation. With a certain frequency, a new vesicle forms from the main compartment and a certain percentage of the main compartment molecules become enclosed in the new vesicle. The percentage values used to determine the initial conditions for the vesicle contents are given in Table 2. Endocytic vesicle formation on cell plasma membranes is a good example of a compartmentalization process.[10,15,16] In the model, the vesicles are formed at a constant rate λ_f (equivalent to a zeroth order reaction, reaction 7, Table 1), and a percentage of the vesicles return back and merge to the main compartment with a rate constant λ_r in a given time period (modeled as a first order reaction, reaction 8). Although some of the rate constants are different (Table 1), molecules in the vesicles continue to react as in the main department (Figure 2), and therefore, the molecular contents of the vesicles can change with time. When a vesicle recycles, its contents are returned to the main compartment. The species C, D, and C:D can only be found in the main compartment, i.e., they are not taken in by the vesicles directly. However, C, D, or C:D can be part of the vesicles if they are bound to the A:B complex (i.e., as A:B:C and A:B:C:D). In the cell signaling network example, the main compartment would correspond to the cytoplasm. Thus, species C, D, and C:D would exist in the cytoplasm but not in the subcellular compartments. The C, D, or C:D molecules that dissociate from the A:B:C and A:B:C:D complexes in the vesicles are added back to the main compartment.

The model employed here can represent many different physical systems. For example, it can be used to model the surface adsorption of molecules. Species A can be the surface molecules that adsorb B molecules by bonding chemically. The adsorption of a B molecule starts a cascade of reactions. If the molecules on a surface have a tendency to form clusters, the compartment formation step can be thought of as the formation

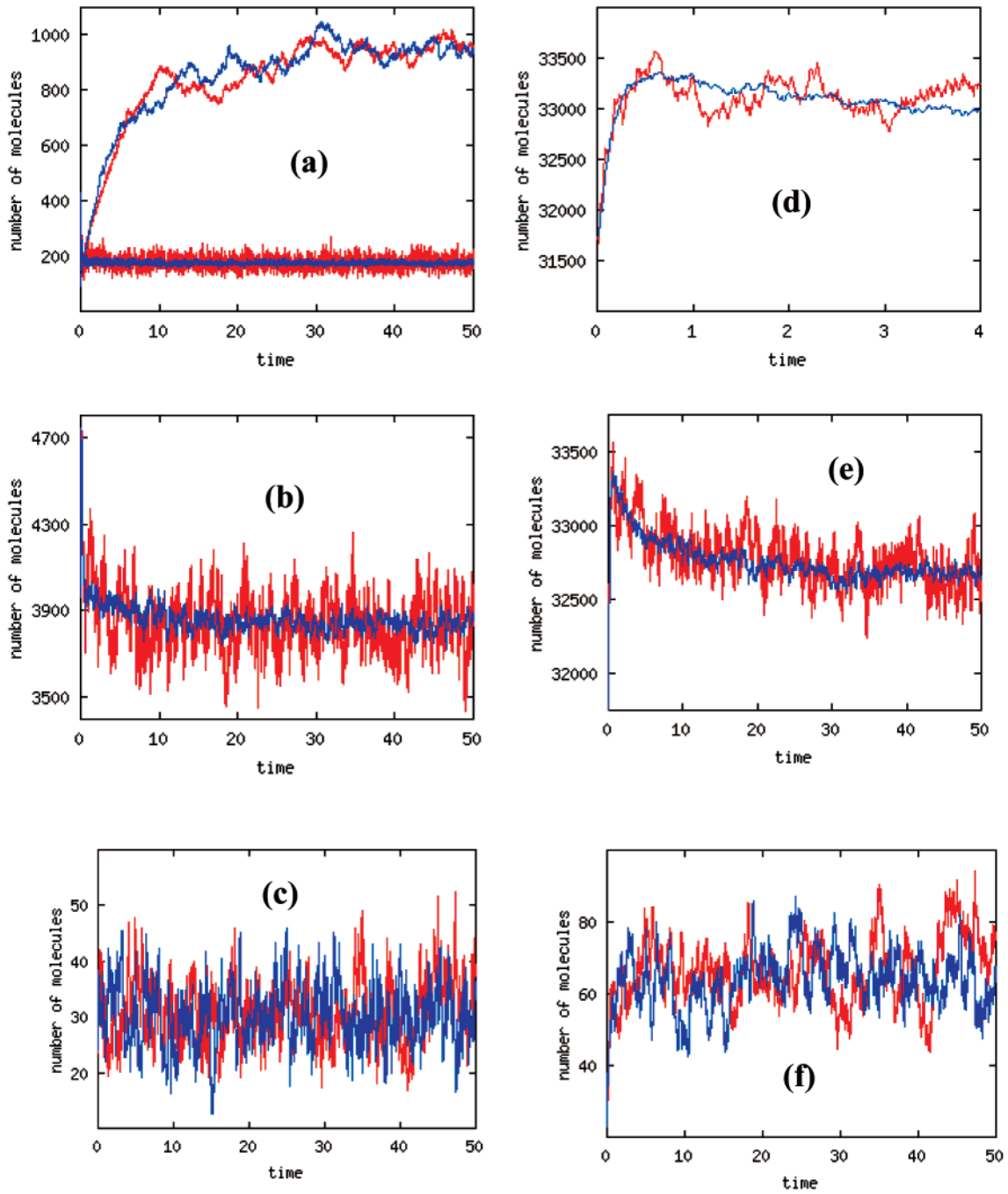

Figure 4. Number distribution of various species in the second model system. Vesicle results are obtained by summing the molecular distribution of all the vesicles. The DMC and the PW-DMC results are shown with blue and red lines, respectively. The horizontal axis shows the simulation time in units of 10^3 seconds. (a) Activated A:B* complex, upper curves are for the main compartment and the bottom ones are for the vesicles. (b) Number of A:B:C molecules in the main compartment. (c) Number of A:B:C molecules in the vesicles. (d) At short time (expanded scale) and (e) number of A:B:C:D molecules in the main compartment. (f) Number of A:B:C:D molecules in the vesicles.

of a new cluster that breaks away from the bulk. If a surface desorption process is studied, the vesicle formation would be the group that breaks away from the surface, and the group could range in size from a single atom to a cluster of molecules.

To investigate the effect of the weighted sampling on steady-state results as well as transient features, the DMC simulations were started from an initial condition that was far away from the steady-state distribution. At the starting point, there are only two vesicles in addition to the main compartment of the system and the number of molecules contained in the vesicles is much smaller than the amount of material in the main compartment. To compare the algorithms, DMC and PW-DMC simulations were run nine times each starting with the same initial distribution but using a different random number seed value. The results were averaged to obtain the expected values and to determine the statistical error bars.[19] The probability of the studied events ranged from near 0 to 86 (note: this is the probability per unit time; these values are multiplied with the time increment so they can be larger than one), and a bin size of 1 was chosen to assign the sampling weights. In other words, events with probabilities less than one had a weight of 1, and the events with probabilities between 65 and 66 had a weight of 66, etc. Figure 3 shows the total number of compartments (= one main compartment + number of vesicles) in the system as a function of time. At the steady state, the expected total number of compartments would be $L = 1 + (\lambda_f/\lambda_r)$. Both

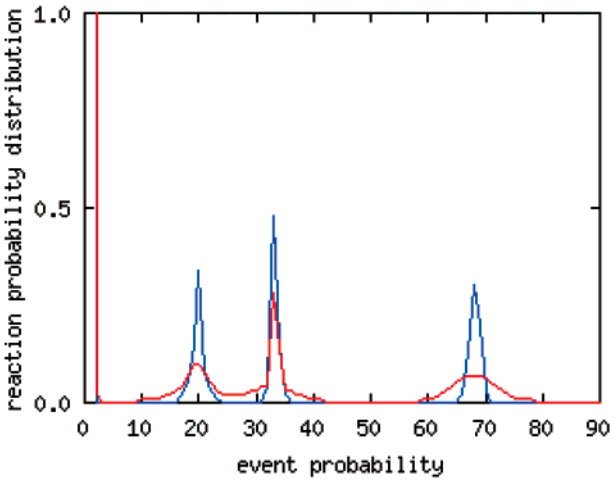

Figure 5. Percentage distribution of the event probabilities (see the text for details): DMC (blue) and PW-DMC (red) lines. The value for the events with probability less than one is 98% for both simulations.

algorithms correctly approach this value at long times. The steady-state distribution is achieved after about 3×10^4 seconds, i.e., after 150 rounds of vesicle formation process. Figure 4 shows the time-dependent distribution of various molecular species. As shown in the figure, the results of the PW-DMC algorithm agree very well with the usual implementation of the DMC method. It should be noted that the simulations were repeated for each method only nine times, actually a small number for such stochastic simulations. For both of the model systems, the number of simulation runs was kept small intentionally to show that the success of the new algorithm does not require a large number of runs. The very good agreement between the algorithms would be expected to become better if additional simulations were performed.[19] The main difference between the results is that the error bars for the PW-DMC method are slightly larger than the DMC method for the distribution in the main compartment for some of the molecular species (Figure 4a−f). This is expected because, in the PW-DMC method, the reactions occur in bundles and this increases the fluctuations in the number of molecules (see eq 6). The increase in the fluctuations of the system is also reflected in the probability distribution of the events. Figure 5 shows the probability distribution of all the events through the simulation. It was calculated by pooling the probabilities of all the events at all steps of the simulations runs. The positions of the peaks in the event probability distributions agree very well. The probability distribution of the events in DMC simulations has sharp peaks. In contrast, the corresponding distribution for the PW-DMC simulations has broader peaks.

We have extended the analysis further by comparing the number of occurrences of the reactions in both sets of simulations. The comparison can be found in Table 1. As the table shows, the largest difference for occurrence of the reactions is less than 3%. In general, the percentage difference is smaller for the reactions that take place more often. Based on the comparison of the PW-DMC and DMC simulation results, it can be confidently stated that, although there is a slight increase in the fluctuations of the system, the approximations due to the use of the PW-DMC method do not have a significant impact on the expected results of the simulations.

The use of the preferential sampling results in a significant gain in the computational efficiency. Each simulation in this study was run for a simulation time of 5×10^4 seconds. The DMC simulations required 12.34 million MC steps to complete, corresponding to 4.05 ms time steps on average. In contrast, only 0.44 million MC steps were required on average for the PW-DMC algorithm. Thus, the PW-DMC simulations needed 28 times fewer Monte Carlo moves than the DMC simulations. The PW-DMC simulations can be made even faster through adjustments in the scale factor. The use of preferential sampling does require the preparation of the probability weight table. The additional expense to compute the weight table is very small (less than 3%), so the overall gain is significant. The model system that we have used to test the PW-DMC algorithm involves a relatively small number of reactions, so it can be studied by using stochastic simulations without the use of smart sampling algorithms on a desktop personal computer. However, the biological network systems that we are actually studying involve thousands of reactions with a wide range of time scales.[15] With the current technology, such large systems cannot be studied in a reasonable time frame by using desktop computers. Thus, smart sampling algorithms, such as the PW-DMC proposed in this work, will be extremely useful in making the investigation of large systems of kinetic rate equations possible without accessing high performance supercomputers. In addition, having access to higher performance computers will enable us to study even larger systems with improved sampling, as well as studying systems with even broader ranges of rate constants.

Discussion

We introduced above a probability weighted sampling algorithm for Dynamic Monte Carlo simulations of reaction kinetics. The new PW-DMC algorithm was tested on two model systems. Comparison of the PW-DMC simulation results with the results of the DMC simulations using the standard algorithm established the success of the new algorithm. Since the weight factors of the PW-DMC algorithm can be chosen in different ways, the algorithm is very versatile and makes it possible to obtain significant gains in computational speed.

The PW-DMC algorithm modifies the probability distribution that is used in generating the Markov chain in the MC simulations. Therefore, the sequence of events that takes place might be affected. However, as our theoretical analysis and the results of simulations using model systems show, the alteration of the probability distribution has only a minimal effect on the calculated results. The ensemble averages of the physical quantities using the PW-DMC method are not affected by the involved approximations and, since it increases the fluctuations in the number of molecules, the statistical errors associated with the new method are expected to be larger than the DMC method. In addition to the increase in the sampling efficiency, the PW-DMC algorithm allows for better sampling of rare events. For systems in which the rare events are important in defining the system dynamics, the new algorithm should actually provide more reliable results by sampling the controlling rare events more often.

We have been studying the signaling network of the EGF receptor.[15,10] This network has thousands of possible reactions with a wide spectrum of reaction rates. Our preliminary results show that the results of the PW-DMC simulations agree very well with the results of the standard implementation of the DMC method. Without the probability-weighted sampling, each run of the EGF receptor signaling network simulation (for a run length of 2 h of simulation time) takes several weeks on a modern personal computer or workstation. This makes the investigation of such networks problematic unless high performance supercomputer resources are utilized. However, the simulation runs can be made more than 60 times faster with

the new PW-DMC algorithm, without any noticeable impacts on the results. This speed gain allows us to complete the simulation runs in less than one day on a personal computer. In the EGFR signaling network, the reactions taking place in different compartments are only loosely connected, which makes the computer code used in the DMC simulation suitable for parallelization. Therefore, the PW-DMC simulations can be made even more efficient in terms of elapsed time by accessing modern high performance massively parallel computers.

Acknowledgment. We thank Bill Cannon and Mihaly Mezei for useful discussions. This research was performed at the W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (PNNL), which is operated for the Department of Energy by Battelle. This research was in part sponsored by the Laboratory Directed Research and Development (LDRD) funds.

References and Notes

(1) *Simulation of Liquids and Solids. Molecular Dynamics and Monte Carlo Methods in Statistical Mechanics*; Ciccotti, G., Frenkel, D., McDonald, I. R., Eds.; North-Holland: Amsterdam, 1987.

(2) Bortz, A. B.; Kalos, M. H.; Lebowitz, J. L. *J. Comput. Phys.* **1975**, *17*, 10.

(3) Gillespie, D. T. *J. Chem. Phys.* **1977**, *81*, 2340.

(4) Gillespie, D. T. *J. Comput. Phys*. **1976**, *22*, 403.

(5) Fichthorn, K. A.; Weinberg, W. H. *J. Chem. Phys.* **1991**, *95*, 1090.

(6) Kang, H. C.; Weinberg, W. H. *J. Chem. Phys.* **1989**, *90*, 2824.

(7) Clark, M. M.; Raff, L. M.; Scott, H. L. *Comput. Phys*. **1996**, *10*, 584.

(8) Jansen, A. P. J.; Lukkien, J. J. *Catal. Today* **1999**, *53*, 259.

(9) Lukkien, J. J.; Segers, J. P. L.; Hilbers, P. A. J.; Gelten, R. J.; Jansen, A. P. J. *Phys. Rev. E* **1998**, *58*, 2598.

(10) (a) *Receptors, Models for Binding, Trafficking, and Signaling*; Lauffenburger, D. A., Linderman, J. J.; Oxford University Press: Oxford, U.K., 1993. (b) *Cell Signaling* Hancock, J. T.; Addison-Wesley Longman: Harlow, England, 1997.

(11) Hassold, G. N.; Holm, E. A. *Comput. Phys*. **1993**, *7*, 97.

(12) Aiello, O. E.; Haas, V. J.; daSilva, M. A. A.; Caliri, A. *Physica A* **2000**, *282*, 546.

(13) Gibson, M. A.; Bruck, J. *J. Phys. Chem. A* **2000**, *104*, 1876.

(14) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys*. **1977**, *23*, 187.

(15) Resat, H.; Miller, J. H.; Dixon, D. A.; Wiley, H. S. In *Currents in Computational Molecular Biology*; El-Mabrouk, N., Lengauer, T., Sankoff, D., Eds.; Les Publications CRM: Montreal, 2001; pp 79−80.

(16) (a) Wiley, H. S.; Burke, P. M. *Traffic* **2001**, *2*, 12. (b) Haugh, J. M.; Lauffenburger, D. A. *J. Theor. Biol.* **1998**, *195*, 187.

(17) Kholodenko, B. N.; Demin, O. V.; Moehren, G.; Hoek, J. B. *J. Biol. Chem*. **1999**, *274*, 30169.

(18) For second or higher order reactions, $N(E)$ is the smallest number of molecules of the species that take part in the event E. For bimolecular reactions of the type $2A \rightarrow B$, the effect of the approximation would be $2w(E)/N(E)$.

(19) In the Dynamic Monte Carlo simulations, the time steps are not uniform. Therefore, the averages and standard deviations of the runs were calculated by interpolating the results of the runs using a uniform time interval. Since the standard error is proportional to the inverse square root of number of runs, increasing the number of simulation runs would decrease the fluctuations in the results.