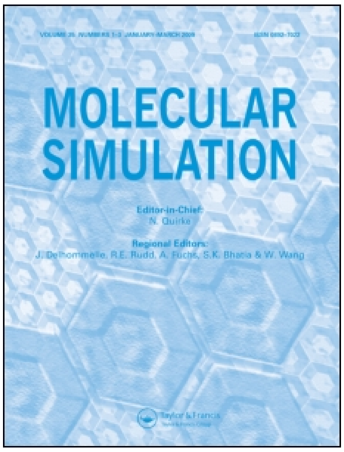

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

A Comparision of Methods for Computing Transition Rates from Molecular Dynamics Simulation

Yuhong Zhang[a]; Richard W. Pastor[a]
[a] Biophysics Laboratory, Center for Biologics Evaluation and Research Food and Drug Administration, Rockville Pike, Bethesda, MD, USA

To cite this Article Zhang, Yuhong and Pastor, Richard W.(1994) 'A Comparision of Methods for Computing Transition Rates from Molecular Dynamics Simulation', Molecular Simulation, 13: 1, 25 — 38
To link to this Article: DOI: 10.1080/08927029408022182
URL: http://dx.doi.org/10.1080/08927029408022182

Molecular Simulation, 1994, Vol. 13, pp. 25-38
Reprints available directly from the publisher
Photocopying permitted by license only

A COMPARISION OF METHODS FOR COMPUTING TRANSITION RATES FROM MOLECULAR DYNAMICS SIMULATION

YUHONG ZHANG and RICHARD W. PASTOR*

Biophysics Laboratory
Center for Biologics Evaluation and Research
Food and Drug Administration
8800 Rockville Pike
Bethesda, MD 20892, USA

(Received June 1993, accepted September 1993)

Correlation function and direct counting methods for the evaluation of the isomerization rates in equilibrium molecular dynamics (MD) simulations are studied for a system of 64 independent particles, each undergoing transitions between two stable states according to a Poisson process. Three different numerical implementations of the number correlation function and two different counting formulas based on different ordering of averages over trajectory and particles are tested. All methods yield correct results for sufficiently long simulations. However, for simulations where a given particle undergoes a small (<5-10) number of transitions, accurate rates are obtained only when the correlation function is normalized *after* averaging over particles; this correlation function is shown to be equivalent to the relaxation function of Brown and Clarke (*J. Chem. Phys.* **92**, 3062, 1990). Similarily, accurate rates are obtained with a counting formula where the average first passage time is first calculated over all particles and inverted, as opposed to calculating rates for each particle and then averaging. These distinctions are shown to be significant when calculating isomerization rates from 500 *ps* MD simulations of *n*-butane. Correction of overcounts is also illustrated for the butane system, where "glassy" or low friction dynamics is evident.

KEY WORDS: isomerization rate, correlation function, relaxation function, counting, molecular dynamics simulation, Poisson process

1 INTRODUCTION

An important issue in computer simulation studies of polymers is the evaluation of rate constants for torsion angle transitions. According to linear-response theory [1, 2], these rate constants can be obtained from either the *correlation functions* of equilibrium fluctuations out of the reactive state, or the *relaxation function* of the decay of a non-equilibrium population [3, 4]. The former method has been used in equilibrium molecular dynamics (MD) [5-7] and Langevin dynamics (LD) [8-10] simulations. In non-equilibrium molecular dynamics (NEMD) simulations the system is prepared in the reactive state and allowed to decay to equilibrium, and therefore relaxation functions are appropriate [11, 12].

---

[*] Author to whom correspondence should be addressed.

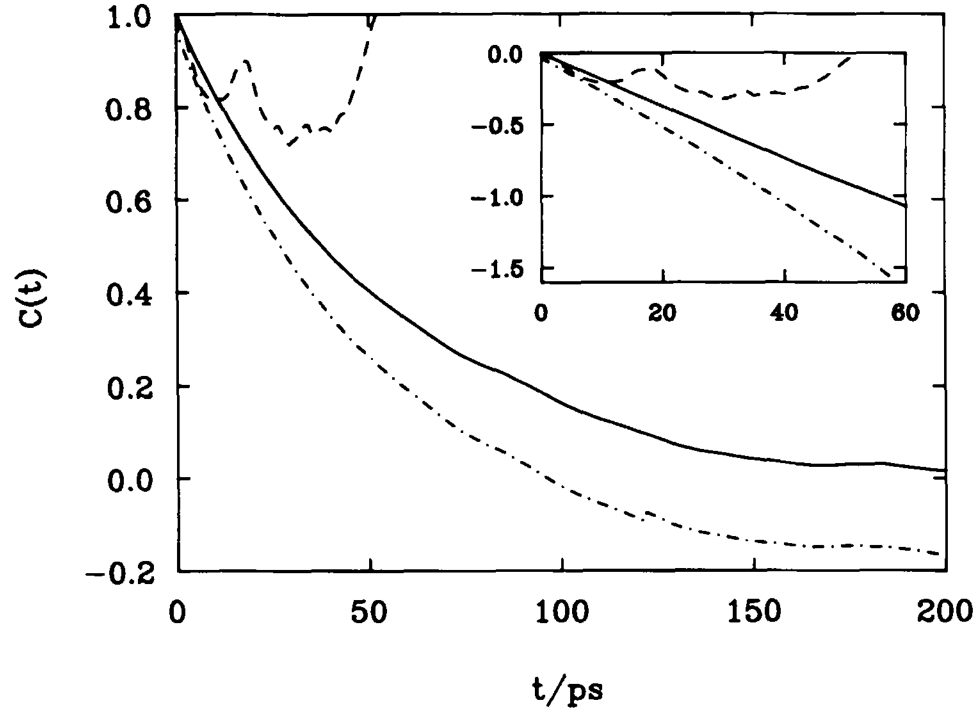

Figure 1 The correlation functions $C_{tot}(t)$ (dashed, Equation 2.8), $C_{ind}(t)$ (dot-dashed, Equation 2.9) and $C_r(t)$ (solid line, Equation 2.10) evaluated from a 500 *ps* MD simulation of liquid *n*-butane with all-atom model (Model 1) at $T = 270\,K$. The inset shows the semi-log plots.

In equilibrium MD simulations, there are different ways to compute the correlation functions, depending on the order of averaging over the trajectory and different particles, and applying the normalization. While these are equivalent in principal, differences might be expected when calculating rates from simulations on the 100 *ps* timescale; i.e., one method may be a better estimator of the true rate in cases where a particular torsional angle (among many) may have undergone only a small number of transitions. Some of these differences have been investigated by Brown and Clarke [7], and are illustrated in Figures 1 and 2, where three correlation functions (defined explicitly in the following section) from two 500 *ps* MD simulations of 64 *n*-butane molecules [13] are compared.

Isomerization rates can also be evaluated from direct counting [9, 10] and related methods using hazard analysis of first passage times [14, 15]. Results from this approach also depend on the order of averaging, and are complicated by the occurrence of recoils and overshoots often found in MD simulations.

Because it is not currently feasible to definitively compare either the correlation function or counting approaches using MD simulations, this paper considers instead a system of N_0 independent particles, each undergoing transitions between two stable states according to a Poisson process. To model the simulations noted above, we set $N_0 = 64$ and the Poisson jump parameters approximately to those of

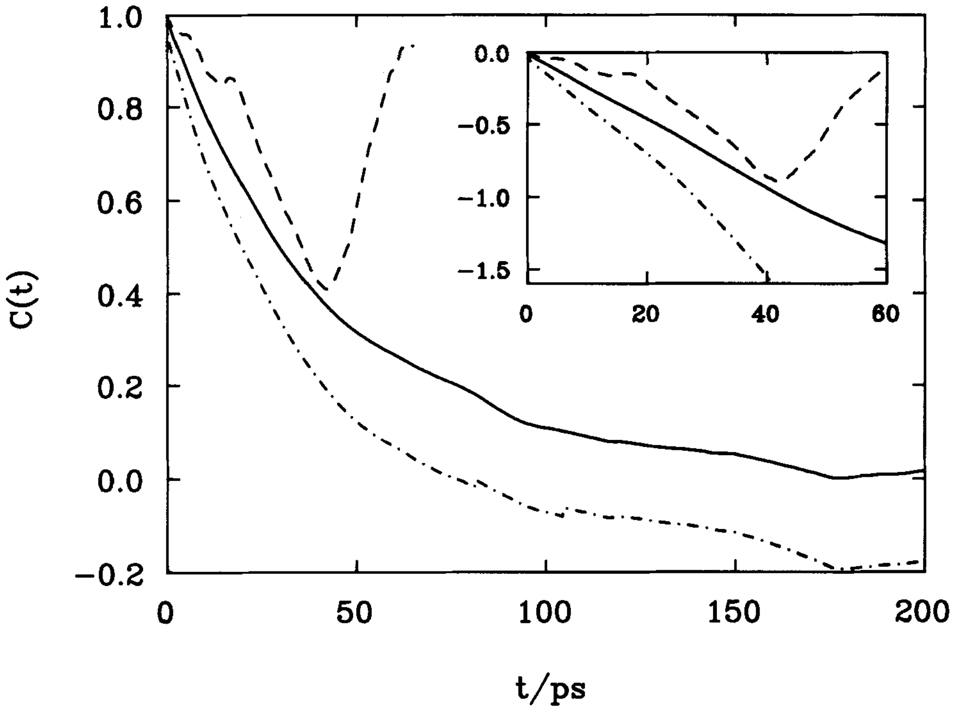

Figure 2 The correlation functions $C_{tot}(t)$ (dashed), $C_{ind}(t)$ (dot-dashed) and $C_r(t)$ (solid line) evaluated from a 500 *ps* MD simulation of liquid *n*-butane with extended-atom model (Model 2) at $T = 270\,K$. The inset shows the semi-log plots.

butane. Time series corresponding to 0.1–50 *ns* MD simulations were then generated, and the means and variances for transition rates were obtained for each method. It is shown that all the methods defined yield correct transition rates for multi-nanosecond simulations, however, the order of averaging is important for simulation times of 500 *ps* or less.

The following section briefly reviews the reaction rate equation for a two-state system, and defines the correlation functions and counting methods. The Poisson simulation method and results for the model butane system are presented and discussed in Section 3. Two MD simulations of *n*-butane based on the all-atom and extended atom models are then analyzed in Section 4.

2 THEORY

2.1 *Reaction Rate Equation*

Consider a stochastic dynamical system consisting of N_0 identical particles. We assume that each particle can jump between two stable states a and b, with k_{ab} the rate constant from a to b, and k_{ba} the reverse. Define the state function for particle i

$$n_{ai} = \begin{cases} 1, & \text{if in state a;} \\ 0, & \text{if in state b.} \end{cases} \tag{2.1}$$

The total number of particles in state a at time t is

$$N_a(t) = \sum_{i=1}^{N_0} n_{ai}(t). \tag{2.2}$$

The state function and the total number of particles in state b are defined in a similar way. The following linear reaction rate equations then hold for the averages of the stochastic variables $\langle N_a(t)\rangle$ and $\langle N_b(t)\rangle$:

$$\frac{d}{dt}\langle N_a(t)\rangle = -k_{ab}\langle N_a(t)\rangle + k_{ba}\langle N_b(t)\rangle, \tag{2.3a}$$

$$\frac{d}{dt}\langle N_b(t)\rangle = k_{ab}\langle N_a(t)\rangle - k_{ba}\langle N_b(t)\rangle. \tag{2.3b}$$

At equilibrium, $\frac{d}{dt}\langle N_a\rangle = \frac{d}{dt}\langle N_b\rangle = 0$, then $k_{ab}\langle N_a\rangle = k_{ba}\langle N_b\rangle$. Combining the preceding equations with the particle conservation condition $N_a(t) + N_b(t) = N_0$, we obtain

$$\langle N_a\rangle = \frac{k_{ba}}{k_{ab} + k_{ba}} N_0. \tag{2.4}$$

Now let us assume that at time $t = 0$ there is a small disturbance that perturbs the system from equilibrium; i.e., $\langle \delta N_a(0)\rangle = (\langle N_a(0)\rangle - \langle N_a\rangle) \neq 0$. It can be then shown from Equation (2.3) that

$$\langle \delta N_a(t)\rangle = \langle \delta N_a(0)\rangle e^{-\lambda t} \tag{2.5}$$

where $\lambda = k_{ab} + k_{ba}$. Equation (2.5) is a relaxation function for this small disturbance. It decays to zero at very long time (as the system rearches equilibrium), and

$$k_{ab} = \lambda\left(1 - \frac{\langle N_a\rangle}{N_0}\right). \tag{2.6}$$

From linear response theory, the correlation function of the particle number associated with the stochastic dynamical equations (2.3) is

$$C(t) = \frac{\langle \delta N_a(t)\delta N_a(0)\rangle}{\langle [\delta N_a]^2\rangle} = e^{-\lambda t}, \tag{2.7}$$

where the average is taken over all equilibrium ensemble samples.

2.2 *Correlation Function*

In MD simulations, the dynamical trajectories of N_0 particles are recorded and the time series $n_a(t)$ is computed from Equation (2.1). The correlation function (2.7) can be implemented in the following three ways:

$$C_{tot}(t) = \frac{\langle N_a(t)N_a(0)\rangle - \langle N_a\rangle^2}{\langle N_a^2\rangle - \langle N_a\rangle^2}, \tag{2.8}$$

$$C_{ind}(t) = \frac{1}{N_0}\sum_{i=1}^{N_0} \frac{\langle n_{ai}(t)n_{ai}(0)\rangle - \langle n_{ai}\rangle^2}{\langle n_{ai}\rangle - \langle n_{ai}\rangle^2}, \tag{2.9}$$

$$C_r(t) = \frac{\sum_{i=1}^{N_0} \langle n_{ai}(t)n_{ai}(0)\rangle - \sum_{i=1}^{N_0} \langle n_{ai}\rangle^2}{\sum_{i=1}^{N_0} \langle n_{ai}\rangle - \sum_{i=1}^{N_0} \langle n_{ai}\rangle^2}. \tag{2.10}$$

The difference between $C_{tot}(t)$ and $C_{ind}(t)$ is the order of the averages over the particle ensemble and the particle trajectories. $C_{tot}(t)$ is obtained by summing the particles first and then averaging over the trajectory. It is the most direct way to apply Equation (2.7), but not necessarily the most efficient. By reversing the above order in calculating $C_{ind}(t)$, one neglects the inter-particle correlation of the state functions.

The difference between $C_{ind}(t)$ and $C_r(t)$ is the order of the normalization of the correlation functions. $C_{ind}(t)$ is obtained by normalizing individual correlation functions first and then averaging over different particles. In calculating $C_r(t)$, one first averages the correlation functions of different particles and then normalizes. We anticipate that the error for $C_{ind}(t)$ is larger than that of $C_r(t)$ for short simulations because the denominators $(\langle n_{ai}^2\rangle - \langle n_{ai}\rangle^2)$ for the individual correlation functions may be very small if the number of transitions is small.

In general, we can expect that for simulations with many transitions per particle

$$C_r(t) \simeq C_{ind}(t) \simeq C_{tot}(t) = e^{-\lambda t}. \tag{2.11}$$

It is interesting to point out that the correlation function $C_r(t)$ is identical to the relaxation function $R(t)$ for a ***equilibrium*** system introduced by Brown and Clarke [7]. Following their suggestion, consider the subsystem of particles in state a in an equilibrium MD simulation at a particular time t_0,

$$N_a(t_0) = \sum_{i=1}^{N_0} n_{ai}(t_0). \tag{2.12}$$

These particles are not necessarily in the same region of space or otherwise divided from the others. After time t_0, the total number of particles in this subsystem will begin to decay to the appropriate equilibrium fraction. The unnormalized relaxation function for this subsystem is

$$\tilde{R}_a(t_0, t_0 + t) = \sum_{i=1}^{N_0} n_{ai}(t_0)n_{ai}(t_0 + t). \tag{2.13}$$

Taking the average over the initial time t_0 and normalizing,

$$R(t) = \frac{\left\langle \sum_{i=1}^{N_0} n_{ai}(t)n_{ai}(0)\right\rangle - \frac{1}{N_0}\langle N_a\rangle^2}{\langle N_a\rangle - \frac{1}{N_0}\langle N_a\rangle^2} \tag{2.14}$$

If we reverse the summation and the average over the trajectory in Equation (2.14), it is clear that the relaxation function $R(t)$ is identical to the correlation function $C_r(t)$. That means the relaxation function method for evaluation the isomerization rates in equilibrium MD simulations introduced by Brown and Clarke is actually a correlation function method. The subscript r in $C_r(t)$ is introduced to signify this equivalence.

2.3 *Counting*

In addition to correlation functions, isomerization rates from simulations can be evaluated from direct counting [9, 10]. This method requires a precise specification of a jump; for example, in "simple counting" a transition is defined as any event where the particle (or torsion angle) crosses over the transition state and passes the minimum of the product well [9, 10].

Let us denote m_i the number of transitions from state a to b for i-th particle and $(\tau_{i1}, \tau_{i2}, \ldots \tau_{im_i})$ the waiting (or first passage) time for each transition. Then the total time for i-th particle in state a is

$$t_i = \sum_{j=1}^{m_i} \tau_{ij} \equiv \langle n_{ai}\rangle T_{run} \tag{2.15}$$

where T_{run} is total simulation time. The average number of transitions from a to b per particle (calculated from the trajectory) is

$$\langle m\rangle = \frac{1}{N_0}\sum_{i=1}^{N_0} m_i \tag{2.16}$$

and the average time spent in state a is

$$\langle t\rangle = \frac{1}{N_0}\sum_{i=1}^{N_0} t_i. \tag{2.17}$$

The rate k_{ab} can be calculated as

$$k_{tg} = \frac{\langle m\rangle}{\langle t\rangle}. \tag{2.18}$$

The rate defined by Equation (2.18) is similar to $C_{tot}(t)$ or $C_r(t)$, in that particle averages $\langle m\rangle$ and $\langle t\rangle$ are taken first. Rates can also be computed by the counting method analogously to C_{ind}:

$$k_{tg} = \frac{1}{N_0}\sum_{i=1}^{N_0} (k_{tg})_i = \frac{1}{N_0}\sum_{i=1}^{N_0} \frac{m_i}{t_i} = \left\langle \frac{m}{t}\right\rangle. \tag{2.19}$$

We can rewrite (2.18) as follows

$$k_{tg} = \frac{\sum_{i=1}^{N_0} m_i}{\sum_{i=1}^{N_0}\sum_{j=1}^{m_i} \tau_{ij}} = \left[\frac{1}{\sum_{i=1}^{N_0} m_i}\sum_{i=1}^{N_0}\sum_{j=1}^{m_i} \tau_{ij}\right]^{-1} = \frac{1}{\langle \tau\rangle}. \tag{2.20}$$

Similarly, Equation (2.19) can be written as

$$k_{tg} = \frac{1}{N_0} \sum_{i=1}^{N_0} \left(\frac{1}{m_i} \sum_{j}^{m_i} \tau_{ij} \right)^{-1} = \frac{1}{N_0} \sum_{i=1}^{N_0} \frac{1}{\langle \tau_i \rangle}. \tag{2.21}$$

From Equation (2.20) and (2.21), it is clear that the method of counting the number of transitions per unit time is equivalent to the method of computing the average first passage time.

3. COMPUTER SIMULATION WITH POISSON PROCESS

3.1 *Method*

The length of a typical MD simulation of a large system is currently several hundred picoseconds so it is not feasible to generate the much longer trajectories (e.g., 100 *ns*) required for comparing the different correlation function approaches using MD.

Fortunately, hundreds of nanoseconds of test data can be rapidly generated by using a Poisson jump model [16], with parameters set approximately equal to the rates obtained from the butane MD simulation. For the present study, the time series n_a for each particle was generated with the following algorithm:

$$n_a(j+1) = n_a(j)\left[1 - H(k_{ab}\Delta t - \xi_j)\right] + \left[1 - n_a(j)\right] H(k_{ba}\Delta t - \xi_j), \tag{3.1}$$

where $n_a(j)$ is the state function at the j-th step, ξ_j is a random number sampled from a uniform distribution in [0, 1], Δt is the timestep and $H(x)$ is the step function

$$H(x) = \begin{cases} 1, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

As an example, consider a particle in state a at the j-th step (i.e., $n_a(j) = 1$). Equation (3.1) specifies that if $\xi_j < k_{ab}\Delta t$, then $n_a(j+1) = 0$ (i.e., the particle jumps to state b); otherwise, $n_a(j+1) = 1$ (it remains in state a). Similarly, for a particle in state b at step j, if $\xi_j < k_{ba}\Delta t$, then $n_a(j+1) = 1$ (the particle jumps to state a); otherwise, $n_a(j+1) = 0$ (it remains in state b).

This algorithm is valid for $k_{ab}\Delta t \ll 1$ and $k_{ba}\Delta t \ll 1$, and directly generates time series analogous to those obtained from an MD simulation. Alternatively, waiting times between jumps can be picked from an exponential distribution [16].

To model the butane MD simulation, we set the two jump rates $k_{ab} = 10\,ns^{-1}$, $k_{ba} = 20\,ns^{-1}$ and timestep $\Delta t = 0.001\,ps$. Different random seeds were assigned to each of the 64 particles, so the transitions of different particles were independent; this ignores inter-particle correlations that may be present in the MD simulation. The equivalent 500 *ps* time series of 64 state functions was generated in less then 1 minute on an Apollo DN10000 workstation.

3.2 *Results*

Figure 3 shows the correlation functions evaluated from single 500 *ps* and 50 *ns* simulations. To further compare the methods, sets of 10 simulations were carried

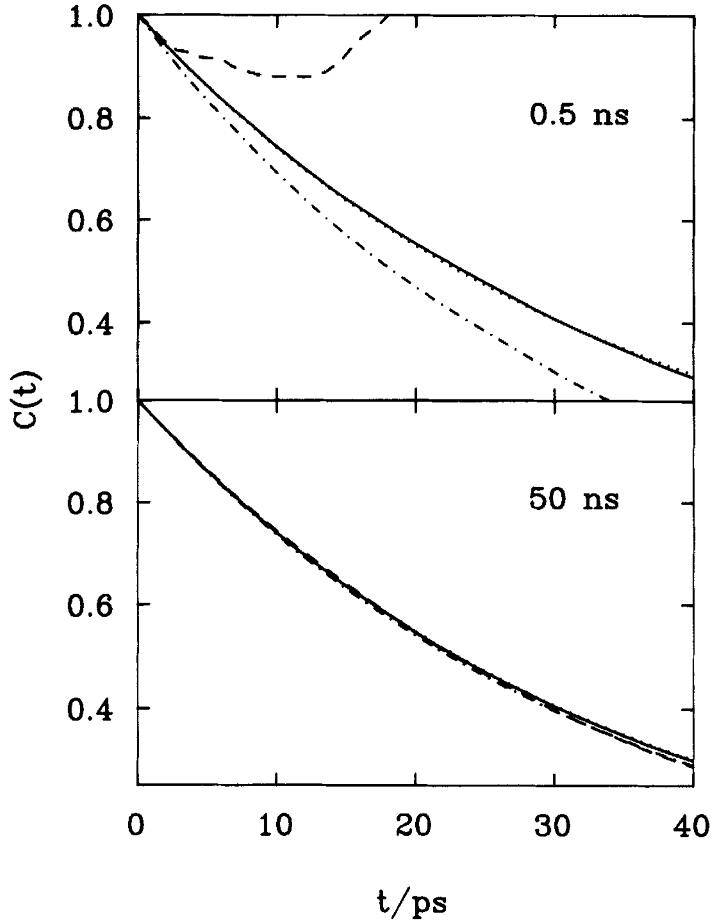

Figure 3 $C_{tot}(t)$ (dashed), $C_{ind}(t)$ (dot-dashed) and $C_r(t)$ (solid) from two Poisson simulations of 0.5 and 50 *ns*. The dotted line shows the exact exponential decay function for the model.

out over the range 0.1 to 20 *ns*. The mean values and standard deviations for k_{ab} are plotted in Figure 4.

In calculating k_{ab}, single exponential fits of the correlation functions were computed over the following two regions: (1) 0–35 *ps* (in which these functions decay to approximately e^{-1}), and (2) 0–5 *ps* (short-time slope). Similar results were obtained in general, with the standard deviation slightly smaller when the short-time slope is used (Figure 4). Exceptions occurred for $C_{tot}(t)$ for the 0.1, 0.5 and 1 *ns* simulations. Because of very high statistical error in these cases, it was not reasonable to use the 0–35 *ps* interval and the data points are not included in Figure 4.

It is clear from Figures 3 and 4 that for simulation times 5 *ns* or greater all three correlation functions yield the correct rate, although the variance is always smaller

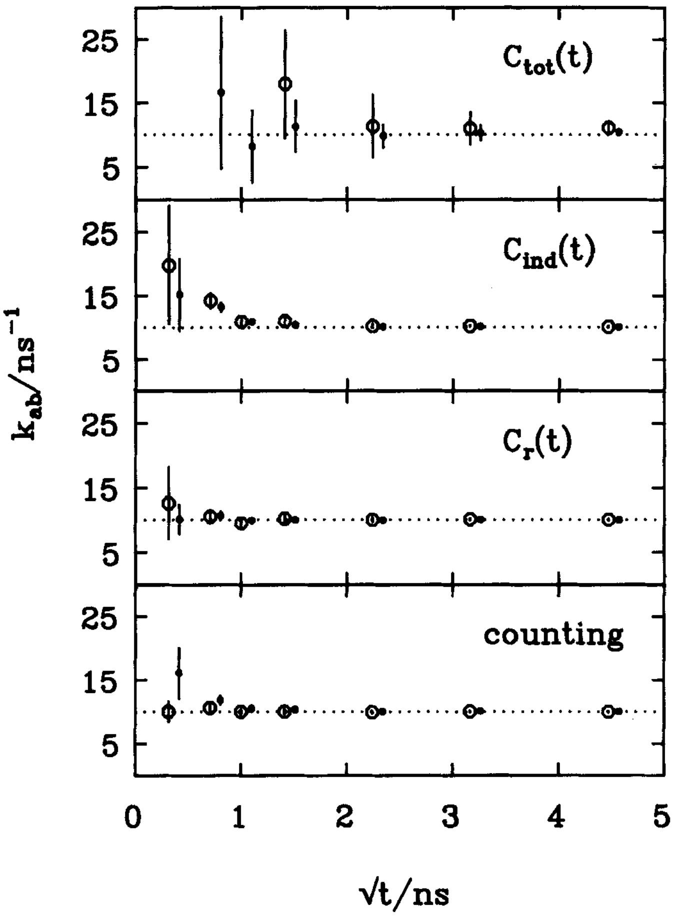

Figure 4 Transition rates k_{ab} and standard deviations from 10 simulations of 0.1, 0.5, 1, 2, 5, 10 and 20 *ns*; dotted line shows exact rate. The simulated rates were obtained from single exponential fits to $C_{tot}(t)$, $C_{ind}(t)$ and $C_r(t)$ over the interval 0–35 *ps* (large circles), and 0–5 *ps* (small circles, slightly offset). The last panel shows the rates calculated by counting method using formula (2.18) (large circles) and formula (2.19) (small circles, slightly offset).

for $C_r(t)$. For example, errors for the 20 *ns* simulations are less than 10% for $C_{tot}(t)$ and less than 1% for $C_{ind}(t)$ and $C_r(t)$.

For the 500 *ps* simulation, however, the $C_r(t)$ is significantly more accurate (~5% error) than the others (40–60% errors). Additionally, it is clear from Figure 4 that $C_{ind}(t)$ systematically overestimates the rate, and the variance in even the short-time slope of $C_{tot}(t)$ is unacceptably high.

For simulations of intermediate length (2 *ns*), the rates from $C_{ind}(t)$ and $R(t)$ are comparable, although $C_{tot}(t)$ still yields poor results.

The last panel in Figure 4 shows the rates obtained by counting. Equation (2.19) systematically overestimates the true rate by approximately the same values as for

$C_{ind}(t)$. The rates computed using Equation (2.18) are accurate at all times, and comparable to those from $C_r(t)$.

3.3 *Discussion*

The Poisson simulations presented here indicate that while the correlation function and counting methods yield correct rates for sufficiently long simulations, $C_r(t)$ and Equation (2.18) are appropriate for shorter (and more typical) simulation times.

The reason for the unacceptably high error in $C_{tot}(t)$ for short simulations is that inter-particle correlations have not yet been averaged to zero. These correlations appear to last a few nanoseconds in the present model; they are due to small sample size and are completely statistical (i.e., the jumps were chosen from independent distributions).

The poor (overestimated) rates associated with $C_{ind}(t)$ in short simulations arise for a different reason. Consider a simulation of total time T_{run}. The average number of transitions μ (a to b and b to a) for each particle is

$$\mu = 2\frac{k_{ab}k_{ba}}{k_{ab} + k_{ba}}T_{run}. \tag{3.3}$$

According to the Poisson distribution, the probability for a particle having m transitions over T_{run} is

$$P(m) = e^{-\mu}\frac{\mu^m}{m!}. \tag{3.4}$$

For example, for $T_{run} = 500\,ps$ and the rate constants specified earlier, each particle has on average 6.67 transitions. However, 6.46 (of $N_0 = 64$) particles have three or fewer transitions. The individual correlation functions $c_i(t)$ of these particles are highly non-exponential, leading to errors in the calculated rate. When the $T_{run} = 5\,ns$, the probability for a given particle having fewer than 40 transitions is small (0.019), leading to well behaved $c_i(t)$ and accurate rates.

This cause of the bias in counting formula (2.19) is also clear. Because $(k_{ab})_i = m_i/t_i$, particles with small t_i (i.e., short time spent in state a) may have very large rates. In long simulations, the probability of particles with small t_i is very small, and the error goes to zero.

It was noted that the slopes evaluated over 0-5 *ps* intervals of the correlation functions yield more accurate rates than those calculated over 0-35 *ps*. This is expected. Because the correlation function from the two-state model is single exponential at all times and the statistical error in normalized correlation functions increases at longer times [17], fitting over a shorter interval is more precise. For MD simulations of polymers, however, the initial (picosecond) decay of the correlation function is non-exponential even for a two-state kinetic model; this "non-phenomenological" behavior arises from the molecular details of the transitions [3, 4]. As a result, short-time estimates of decay constants can be misleading, and it is more reliable to use a larger interval (e.g., $1 \geq C(t) \geq e^{-1}$).

4 ANALYSIS OF BUTANE SIMULATIONS

To test these methods on a molecular system, we now return to the MD simulations of liquid *n*-butane at 270 K noted in Section 1. Results from two models are presented: Model 1 (all-atom, using the force field PARM22 [18] of CHARMM [19]); and Model 2 (extended-atom, using PARM19 [20]). Each simulation contained 64 molecules in a cubic box of 22 Å with periodic boundary conditions, and was carried out with constant particle number, volume and energy, and timestep 0.001 *ps* using the Verlet algorithm [21]. The analysis described here is based on 500 *ps* trajectories following equilibration. Rates were obtained from single exponential fits to the correlation functions in the interval $1 \geq C(t) \geq e^{-1}$. Additional details are presented in reference 13.

The trans-to-gauche isomerization rates k_{tg} computed from $C_{ind}(t)$ and $C_r(t)$ are $4.89\,ns^{-1}$ and $3.29\,ns^{-1}$ for Model 1 (Figure 1); for Model 2 (Figure 2) they are $7.12\,ns^{-1}$ and $4.32\,ns^{-1}$, respectively. Hence, as already shown from the Poisson simulations, there are significant differences between the methods on the 500 *ps* timescale, and the rates obtained from $C_r(t)$ are more accurate. From the error formula of reference [17], the statistical uncertainty in these values is approximately 5%.

To evaluate the rate by counting, we slightly modify Equation (2.20):

$$k_{tg} = \frac{\langle m \rangle}{2\langle t \rangle}, \tag{4.1}$$

where $\langle t \rangle$ is the average time spent in the trans state, and the factor of 1/2 takes into account that a butane in trans can isomerize to either gauche plus or gauche minus. For the present 500 *ps* simulations, $\langle t \rangle = 317.223\,ps$ (Model 1) and $313.208\,ps$ (Model 2). Applying "simple counting" (defined in Section 2.3), $\langle m \rangle = 2.938$ (Model 1) and 6.625 (Model 2), yielding $k_{tg} = 4.63\,ns^{-1}$ (Model 1) and $10.58\,ns^{-1}$ (Model 2). These values are significantly larger than those obtained from $C_r(t)$. The reason is overcounting of the isomerizations.

For molecules exhibiting either "glassy" or low friction dynamics, simple counting, which includes recoils and overshoots, leads to an overestimate of the true rate; i.e., such events do not contribute to the exponential decay of the reactant population, and therefore are incorrectly labeled transitions. Figures 5 and 6 show portions of butane trajectories illustrating true transitions, recoils and overshoots for both models; low friction behavior from a different butane simulation is shown in Figure 8 of reference 7. As might be expected for an extended atom parameter set, Model 2 shows much lower friction behavior than does Model 1 (i.e., the hydrogens add friction to the internal motion).

A straightforward way to correct for overcounts is to stipulate that for an event to qualify as a transition, the torsion angle must remain in the reactant state for a minimum time (t_{min}) after the jump. For example, if $t_{min} = 0.5\,ps$ (approximately 2.5 oscillations), there are 42 and 203 overcounts for Models 1 and 2, respectively. Subtracting these overcounts results in $k_{tg} = 3.60\,ns^{-1}$ (Model 1) and $5.51\,ns^{-1}$ (Model 2); for $t_{min} = 1.0\,ps^{-1}$, $k_{tg} = 3.37\,ns^{-1}$ (Model 1) and $4.94\,ns^{-1}$ (Model 2). Both estimates for Model 1 agree well with the value $3.29\,ns^{-1}$ obtained from $C_r(t)$. To obtain an accurate rate for Model 2, however, t_{min} must be increased to $2\,ps$; this yields in 244 overcounts or $k_{tg} = 4.49\,ns^{-1}$, in good agreement with

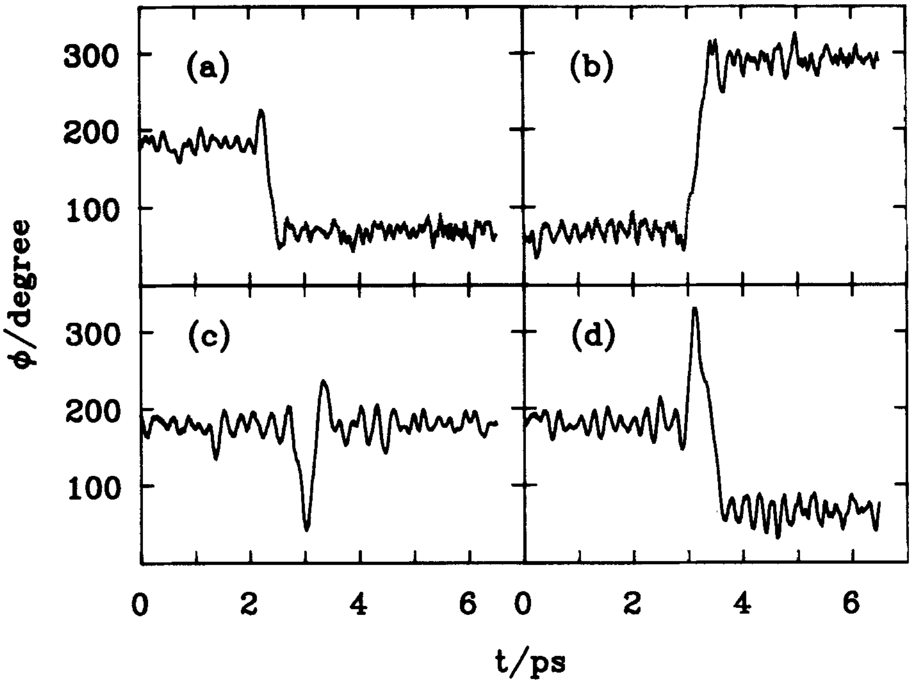

Figure 5 Sample *n*-butane (Model 1, all-atom) torsional angle trajectories illustrating low friction dynamics: (a) a successful transition from trans to gauche minus; (b) a transition from gauche minus to gauche plus (an overshoot through trans); (c) a recoil from gauche minus (no transition); (d) a successful transition from trans to gauche minus, with intervening recoil, and overshoot. A minimum waiting time of 1 *ps* is assumed (see text).

$4.32\,ns^{-1}$ from $C_r(t)$. (In contrast, if t_{min} is increased from 1 *ps* to 2 *ps* for Model 1, only 1 additional overcount is found.)

It was shown earlier [10] for a Langevin bistable oscillator that the appropriate value of t_{min} approximately equals the kinetic energy relaxation time. Although it is clearly difficult to specify *a priori* these criteria for arbitrary systems, it is advisable to scrutinize any rate calculated only by simple counting.

Transition state sampling has been used to calculate isomerization rates of a single butane in model carbon tetrachloride [5] and a bistable diatomic in a Lennard-Jones fluid [6]. In this method, the molecule is placed at its transition state and simulated with different initial velocities. The reactive flux (related to the derivative of the correlation function $\langle n_a(t)n_a(0)\rangle$) is evaluated, and the rate is obtained from the plateau value [3, 4]. Because this plateau value is reached quickly (e.g., 1 to 2 *ps* for butane), it is possible to carry out many short simulations and average. This is clearly the method of choice for determining the transition rate of a single molecule in a solvent of different molecules. However, it is probably comparable to the methods described above when simulating neat solutions. For example, if a single butane were placed at the transition state and 2 *ps* trajectories were carried out, 250 samples would be obtained with 500 *ps* total simulation time. If we add

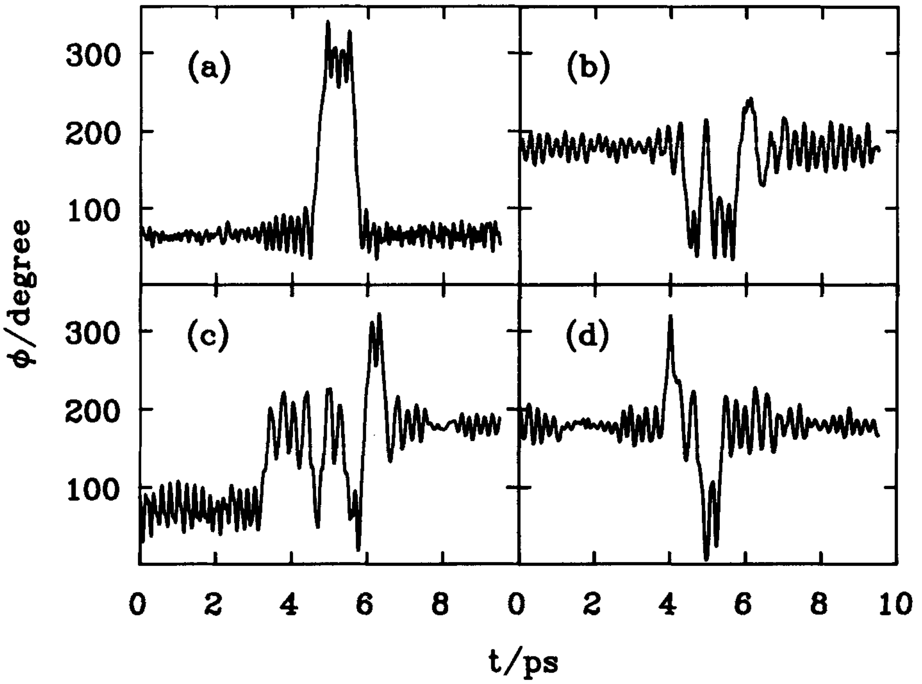

Figure 6 Sample *n*-butane (Model 2, extended-atom) torsional angle trajectories illustrating very low friction dynamics. Assuming a minimum waiting time of 2 *ps* (see text), only panel c shows a successful transition from gauche plus to trans.

the total number of trans-to-gauche and gauche-to-trans crossings of the barrier in the present 500 *ps* simulation, there were 325 transition state samplings for Model 1. It is obviously possible to generate a larger number of samples with transition state sampling by placing more than one molecule at the barrier; however, this invites the danger that the system may be too far from equilibrium for the assumptions of linear response theory to be valid. (NEMD simulations have the same potential problem.) Finally, an added advantage of long equilibrium trajectories is that properties with longer correlation times (e.g., rotational relaxation) can be studied with the same simulation.

5 CONCLUSIONS

Direct counting is a reliable method for calculating rates provided one can identify and then subtract all the recoils and overshoots over the potential barrier. Such overcounts are not negligible in systems where the effective damping constants are small. Recoils and overshoots are not of concern, however, when correlation functions are used; i.e., they make a small contribution to the relaxation at very short times ($< 0.5\,ps$ for the present simulations). For the methods studied in detail here, the correlation function $C_r(t)$ is the best for the evaluation of isomerization rates for equilibrium MD simulations in the timescale of 100 *ps*. It is simple, straightforward and accurate.

Acknowledgements

We thank Attila Szabo, George Weiss and Peter Munson for helpful discussions.

References

[1] R. Zwanzig, "Linear Response Theory" in *Molecular Fluids* (Les Houches 1973b), edited by R. Balian and G. Weill (Gordon and Breach), pp 1-36 (1976).

[2] R. Kubo, "The Fluctuation-dissipation theorem", *Rep. Progr. Phys.*, **29**, 255 (1966).

[3] D. Chandler, "Statistical mechanics of isomerization dynamics in liquids and the transition state approximation", *J. Chem. Phys.*, **68**, 2959 (1978).

[4] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.

[5] R.O. Rosenberg, B.J. Berne, and D. Chandler, "Isomerization dynamics in liquids by molecular dynamics", *Chem. Phys. Lett.*, **75**, 162 (1980).

[6] J.E. Straub, M. Borkovec, and B.J. Berne, "Molecular dynamics study of an isomerizing diatomic in a Lennard-Jones Fluid", *J. Chem. Phys.*, **89**, 4833 (1988).

[7] D. Brown, and J.H.R. Clarke, "A direct method of studying reaction rates by equilibrium molecular dynamics: application to the kinetics of isomerization in liquid *n*-butane", *J. Chem. Phys.*, **92**, 3062 (1990).

[8] R.M. Levy, M. Karplus, and J.A. McCammon, "Diffusive Langevin Dynamics of Model Alkanes", *Chem. Phys. Lett.*, **65**, 4 (1979).

[9] R.P. Pastor, and M. Karplus, "Inertial effects in butane stochastic dynamics", *J. Chem. Phys.*, **91**, 211 (1989).

[10] R.J. Loncharich, B.R. Brooks, and R.W. Pastor, "Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of *N*-Acetylalanyl-*N'*-Methylamide", *Biopolymers*, **32**, 523 (1992).

[11] D.J. Evans and G.P. Morriss, 1990, *Statistical Mechanics of Nonequilibrium Liquids*, Academic Press, 1990.

[12] R. Edberg, D.J. Evans and G.P. Morriss, "Conformational Kinetics in liquids butane by nonequilibrium molecular dynamics", *J. Chem. Phys.*, **87**, 5700 (1987).

[13] Y. Zhang, R.M. Venable and R.W. Pastor, *in preparation*.

[14] E. Helfand, "Brownian dynamics study of transitions in a polymer chain of bistable oscillators", *J. Chem. Phys.*, **69**, 1010 (1978).

[15] E. Helfand, Z.R. Wasserman, and T.A. Weber, "Brownian dynamics study of polymer conformational transitions", *Macromolecules*, **13**, 526 (1980).

[16] P.G. Hoel, S.C. Port and C.J. Stone, *Introduction to Stochastic Process*, Houghton Mifflin Company, 1972.

[17] R. Zwanzig and N.K. Ailawadi, "Statistical Error Due to Finite Time Averaging in Computer Experiments", *Phys. Rev.*, **182**, 280 (1969).

[18] A. MacKerell, M. Schlenkrick, J. Brickmann and M. Karplus, unpublished.

[19] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, "CHARMM, A Program for Macromolecular Energy, Minimization and Dynamics Calculations", *J. Comp. Chem.*, **4**, 187, (1983).

[20] W.E. Reiher, Ph.D thesis, Harvard University, Cambridge, MA, 1985.

[21] L. Verlet, "Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules", *Phys. Rev.*, **159**, 98, (1967).