Master equation approach to folding kinetics of lattice polymers based on conformation networks

Yu-Pin Luo and Ming-Chang Huanga
Center for Nonlinear and Complex Systems and Department of Physics, Chung-Yuan Christian University, 320 Chungli, Taiwan

Jinn-Wen Wu
Department of Applied Mathematics, Chung-Yuan Christian University, 320 Chungli, Taiwan

Tsong-Ming Liaw and Simon C. Lin
Grid Computing Centre, Academia Sinica, Taipei, Taiwan and Institute of Physics, Academia Sinica, 11529 Taipei, Taiwan

(Received 24 January 2007; accepted 2 February 2007; published online 4 April 2007)

Based on the master equation with the inherent structure of conformation network, the authors investigate some important issues in the folding kinetics of lattice polymers. First, the topologies of conformation networks are discussed. Moreover, a new scheme of implementing Metropolis algorithm, which fulfills the condition of detailed balance, is proposed. Then, upon incorporating this new scheme into the geometric structure of conformation network the authors provide a theorem which can be used to place an upper bound on relaxation time. To effectively identify the kinetic traps of folding, the authors also introduce a new quantity, which is employed from the continuous time Monte Carlo method, called rigidity factor. Throughout the discussions, the authors analyze the results for different move sets to demonstrate the methods and to study the features of the kinetics of folding. © 2007 American Institute of Physics. [DOI: 10.1063/1.2711816]

I. INTRODUCTION

The problem how proteins fold, in milliseconds to seconds, into unique and stable structures with definite biological functions has recently become intriguing to the biophysicists.1 The kinetic feature of such problem amounts to the characteristics of the folding paths. Subject to this, considerable progress have been achieved through numerical studies of lattice heteropolymers in two or three dimensions.2–4 Although oversimplified, characteristic features obtained from the simulations, such as folding funnel,5,6 folding bottleneck,2 and kinetic traps,3,7 have provided much insights to the kinetic process. However, there exists some suspicions about the simulation method8–10 and few ambiguities in the move sets adopted for the simulations.3,10 For the former, question about implementing the kinetic Monte Carlo algorithm was raised and some implementations were shown to violate the condition of detailed balance.7 For the latter, Chan and Dill8 and Hoang and Cieplak10 stressed the strong dependence of the folding landscape on the choice of move sets. Consequently, further clarification and improvement for the methodology remain essential. In Refs. 9 and 10, proposals of refining the Monte Carlo algorithm were discussed. In addition, move steps other than the conventional ones, such as the snake move,10 have also become of interest.

The master equation approach was introduced by Cieplak et al.7 to the study of folding kinetics of lattice proteins. In this work, we attempt a systematic analysis for the effect of different move sets on the kinematics of foldings based on the master equation approach in the frame of conformation network.11,12 For the conformation network associated with a lattice chain, the nodes are all possible self-avoiding conformations, while the edge distribution among nodes depends on move set. A move set consists of a number of elementary move steps, and there exists an edge between two nodes if a move step can transfer from one to the other. Thus, the choice of move sets in general renders the edge distribution distinct. Based on the inherent structure of conformation network, first we give a straightforward implementation of Metropolis algorithm.13 The new scheme is demonstrated to be accurate and efficient by explicit simulations. Then, the master equation with the input of transition rates obtained from the new proposal is used to study the relaxation times of different move sets. Following the theorem shown in the Appendix, we are able to place an upper bound, which varies with move set and temperature, on relaxation time. Moreover, we employ Gillespie algorithm of continuous time Monte Carlo simulations14 to introduce rigidity factor to a node of conformation network.9 By comparing the rigidity factors in different conformational distances to the native state, we are able to locate the bottleneck of folding.2 We further demonstrate that the folding kinetics can be exhibited clearly by assigning the rigidity factors to conformation distances. This paper is organized as follows: The lattice sequence, the Hamiltonian used to assign energies to confirmations, the move sets, and the structures of the networks associated with different move sets are described in Sec. II. Different implementations of Monte Carlo algorithm including our new proposal and their numerical experiments

---

*Electronic mail: ming@phys.cycu.edu.tw*
are presented in Sec. III. The method of calculating the upper bound of relaxation time and the calculation results for different move sets at various temperatures are given in Sec. IV. The rigidity factor of a node and the locations of the bottleneck of folding are given in Sec. V. Finally, we summarize the results in Sec. VI.

II. MODEL, MOVE SET, AND NETWORK

Based on the considerations given in Ref. 15, we consider the $H$-$P$ (hydrophobic-hydrophilic) model on a two-dimensional square lattice with the "proteinlike" sequence of 16 monomers specified as $PHPPPHPPPPPHP$. All possible conformations can be constructed by self-avoiding random walks with the restriction that two monomers cannot occupy the same lattice site. The energy of a conformation is assigned according to the Hamiltonian

$$H = \sum_{i<j} E_{\sigma_i\sigma_j} \Delta(i,j).$$

Here $E_{\sigma_i\sigma_j}$ represents the contact energy between residues $\sigma_i$ and $\sigma_j$ with $\sigma=H$ or $P$, and we have $\Delta(i,j)=1$ for two nearest-neighbor sites $i$ and $j$ but not adjacent along the chain, otherwise $\Delta(i,j)=0$. By assigning the contact energies with $E_{H,H}=-3.3$, $E_{H,P}=-2$, and $E_{P,P}=-1$, we obtain the energy of the native state as $E_N=-19.2$. Moreover, the average energy gap, defined as the average energy difference between the lowest ten excited states and the native one, is $\Delta E_N=1.97$. The specific-heat curve $C_N(T)$ versus $T$ are shown in Fig. 1 with the inset for the native conformation. In the specific-heat curve, there is a peak representing the phase transition form the molten globule to the native state, together with a shoulder at a higher temperature representing the transition from random coil to molten globule. The folding temperature $T_f$, defined as the temperature at which the probability of the native conformation is 1/2, is at $T_f=0.50$ which is near the specific-heat peak. We are concerned with move sets consisting of four typical elementary moves, the end flip (ef), corner shift (cs), crankshaft (cr), and the rigid rotation (rr). In this work, we focus the study on the move sets $MS_1$, $MS_2$, and $MS_3$, defined as follows. The conventional move set $MS_1$ consists of ef, cs, and cr based on locality. However, in general, the ergodicity cannot be satisfied by $MS_1$. In two dimensions, it prohibits the reaching of one conformation from the others for 16 monomers, and the number of such conformations increases rapidly for more monomers and/or dimensions. The problem can be remedied by involving moves of $rr$ type which have been realized in some simple diffusive motions for groups of monomers. While ef itself can be viewed as short-scale rigid rotation, an ergodic move set, say, $MS_2$, can be achieved by simply combining ef with rr. Finally, the ergodic move set $MS_3$ contains all the moves of four types.

The total number of conformations for a lattice chain of 16 monomers is $K=802075$, which unambiguously provides the total node number, while it remains crucial to determine the edge distributions according to various choices of move sets. However, the corresponding geometric properties of the conformation networks for the monomer numbers $n=16$ has been thoroughly reported. In the following, we give the results for the case of $n=16$ with the specified native state as the target of foldings. These results may provide a solid foundation for understanding qualitatively the latter results.

The number of edges $k_i$ associated with a node $i$ is defined as the number of nodes connecting to the node $i$ by means of elementary moves. The corresponding mean values per node $\langle k \rangle$, defined as

$$\langle k \rangle = \frac{1}{K} \sum_{i=1}^{K} k_i,$$

are 9.2, 20.7, and 26.3 for $MS_1$, $MS_2$, and $MS_3$, respectively. The $\langle k \rangle$ value for $MS_2$ or $MS_3$ is about twice more than the value for $MS_1$. Note that a larger $\langle k \rangle$ value provides more thorough-way accessibility to the native state and reduces the possibility of being trapped in local minimum. Another important quantity is the minimal edge number $l_i$ for a node $i$ to connect with the native state of the model sequence. By defining the characteristic distance of foldings $\langle l \rangle$ as

$$\langle l \rangle = \frac{1}{K} \sum_{i=1}^{K} l_i,$$

we obtain $\langle l \rangle=26.4$ for $MS_1$ and 8.4 for both $MS_2$ and $MS_3$. The $\langle l \rangle$ value for $MS_2$ or $MS_3$ is about one-third of that for $MS_1$. Thus, the first passage time, defined as the required time from a stretched chain to the first arrival of the native state, for the simulations based on $MS_2$ or $MS_3$ can be expected to be much shorter than that for $MS_1$. More detailed informations about the values of $k_i$ and $l_i$ are provided by the distribution functions $P(k)$ and $P(l)$, where $P(k)$ is the probability for a random chosen node to have $k$ edges and $P(l)$ is the probability for a random chosen node to connect to the native state with the minimal edge number $l$. The plots of $P(k)$ versus $k$ and $P(l)$ versus $l$ for different move sets are shown in Fig. 2. The results indicate that a few nodes possess great amounts of edges for $MS_2$ and $MS_3$; accordingly, the distance from a node to the native state is obviously reduced in relation to the case where $MS_1$ is adopted. The results
shown in Fig. 2 also assure us that the kinetic features of MS$_3$ and MS$_1$ are similar but are drastically different from that of MS$_1$.

**III. MONTE CARLO IMPLEMENTATIONS**

The above geometric properties of the conformation networks for different move sets are closely related to the folding dynamics of the model sequence at high temperatures. In practice, the properties of the folding paths at finite temperatures are often extracted from the results of numerical simulations for which, the simulation method and move set are essential. In this section, different implementations of the Metropolis algorithm, including our new proposal, are discussed. To verify the reliability and efficiency of various methods, we also report the simulation results for different simulation methods and move sets.

Consider the probability of finding the sequence in the node $\alpha$, denoted as $P_\alpha(t)$. The time evolution of $P_\alpha(t)$ is governed by the master equation

$$\frac{\partial P_\alpha(t)}{\partial t} = \sum_{\beta, \gamma} [w_{\alpha\beta}P_\beta(t) - w_{\beta\alpha}P_\alpha(t)],$$

where $w_{\alpha\beta}$ is the transition rate from the node $\beta$ to $\alpha$ with $1 \leq \alpha, \beta \leq K$. By referring the transpose of the row $(P_1, \ldots, P_K)$ as the vector $\mathbf{P}$, we can put the master equation in a matrix form

$$\frac{\partial}{\partial t} \mathbf{P}(t) = \mathbf{M} \cdot \mathbf{P}(t),$$

where the elements of $\mathbf{M}$ are

$$m_{\alpha\beta} = w_{\alpha\beta} - \delta_{\alpha\beta} \sum_{\gamma} w_{\gamma\alpha},$$

with the properties, $m_{\alpha\beta} \geq 0$, $m_{\alpha\alpha} < 0$, and $\sum_{\beta} m_{\alpha\beta} = 0$.

The equilibrium probability distribution $\mathbf{P}_{eq}$, which satisfies the condition $\mathbf{M} \cdot \mathbf{P}_{eq} = 0$, is a solution of Eq. (4). The condition is usually replaced by a more restricted condition, called the condition of detailed balance,

$$w_{\beta\alpha}P_{eq, \alpha} = w_{\alpha\beta}P_{eq, \beta},$$

in Monte Carlo simulations. To determine the transition rate $w_{\beta\alpha}$ from the above condition, we first notice that the rate can be expressed as the product of two factors,

$$w_{\beta\alpha} = p_{\beta\alpha}\bar{w}_{\beta\alpha},$$

where $p_{\beta\alpha}$ is the probability of updating $\alpha$ by $\beta$, and $\bar{w}_{\beta\alpha}$ is the acceptance rate of the update. For the acceptance ratio, defined as $R_{\beta\alpha} = \bar{w}_{\beta\alpha}/\bar{w}_{\alpha\beta}$, taking the form of

$$R_{\beta\alpha}^{(0)} = \exp[-(E_\beta - E_\alpha)/T],$$

Eq. (7) is satisfied by

$$\bar{w}_{\beta\alpha}^{(0)} = \frac{R_{\beta\alpha}^{(0)}}{1 + R_{\beta\alpha}^{(0)}},$$

provided

$$p_{\beta\alpha}^{(0)} = p_{\alpha\beta}^{(0)}.$$  

Here, the Boltzmann constant is set to unity and $T$ is the temperature.

A trivial setting, $p_{\beta\alpha}^{(0)} = 1$ for a random choice of connected node $\beta$, fulfills the symmetry condition of Eq. (11), but the corresponding $\mathbf{P}(t)$ fails to converge towards thermal equilibrium. In fact, the setting violates the condition of detailed balance. Collet proposed an elegant solution which reflects the condition of Eq. (7). Here, for illustration of the solution we determine the form of $p_{\beta\alpha}$ by considering the sequence of 16 monomers with the move set MS$_1$, which contains all the moves of rr, cs, ef, and cr. Consider a move pertaining to MS$_3$ to be contained either in MS$_1$ or in MS$_2$. Suppose the corresponding probabilities are then denoted as $r$ and $1 - r$, respectively. Note that the overlap between MS$_1$ and MS$_2$ caused by the move ef is a small portion of the edges, hence, is neglected. For the conformation network shown in Sec. II, the maximal edge number of a node is 18 for MS$_1$ and 28 for MS$_2$. Then, after proper normalization we set

$$p_{\beta\alpha}^{(0)S_1} = \frac{r}{18} \delta_{\beta\alpha}^{S_1},$$

$$p_{\beta\alpha}^{(0)S_2} = \frac{1 - r}{28} \delta_{\beta\alpha}^{S_2},$$

where $\delta_{\beta\alpha}^{S_1}(\delta_{\beta\alpha}^{S_2}) = 1$ if $\alpha$ and $\beta$ are connected by a move of MS$_1$(MS$_2$), otherwise $\delta_{\beta\alpha}^{S_1}(\delta_{\beta\alpha}^{S_2}) = 0$. Note that by averaging over all nodes, the probability of connecting two nodes is 0.35 by MS$_1$ and 0.65 by MS$_2$. However, the affinity for MS$_1$ should be larger than that for MS$_2$ because of the locality of MS$_1$. 

**FIG. 2.** (a) The distribution function of the edge number associated with a node $P(k)$ vs the edge number $k$ and (b) the distribution function of the distance $l$ from a conformation to the native conformation $P(l)$ vs the distance $l$ for the move sets MS$_1$ (circles), MS$_2$ (triangles), and MS$_3$ (black dots).
MS$_1$. By assuming that one factor compensates for the other, we set $p^{(0)}_{ba} = p^{(0)}_{ab}$ and this leads to $r = 9/23$. Then, the probability to update $\alpha \rightarrow \beta$ is

$$p^{(0)}_{ba} = \frac{1}{46} \delta_{ba},$$

(14)

with $\delta_{ba} = \delta_{ba}^1 + \delta_{ba}^2$, and the probability to update the node $\alpha$ becomes

$$\sum_{\beta} p^{(0)}_{ba} = \frac{k_\alpha}{46},$$

(15)

where $k_\alpha$ is the edge number of the node $\alpha$.

The method indeed yields a convergent evolution to $P_{eq}$ for $P(t)$. However, there is an alternative which is more straightforward and shown to approach the thermal equilibrium distribution very efficiently. Note that Eq. (11) may not be required provided the acceptance rate of Eq. (10) is modified properly. Consider the update $\alpha \rightarrow \beta$ for which the new node $\beta$ is chosen randomly from the $k_\alpha$ connected nodes of $\alpha$. This gives $p_{ba} = \delta_{ba}/k_\alpha$, where $\delta_{ba}=1$ if $\alpha$ and $\beta$ are connected by an edge, otherwise $\delta_{ba}=0$. Then, from Eqs. (7) and (8) the acceptance ratio becomes

$$R_{ba} = \delta_{ba} C_{ba} \exp[-(E_\beta - E_\alpha)/T],$$

(16)

where $C_{ba} = k_a/k_\beta$ and $k_\beta$ is the edge number of $\beta$. Consequently, the acceptance rate is

$$\bar{w}_{ba} = \frac{R_{ba}}{1 + R_{ba}},$$

(17)

which depends both on the energy difference and on the edge numbers of two nodes.

Numerical simulations with different implementations and move sets have been performed. We generate 100 Monte Carlo trajectories at $T = 1.74$ for different cases. Each trajectory begins at the completely extended chain and consists of $3 \times 10^9$ steps. For a trajectory, we measure the 1-norm deviation of probability distribution from the equilibrium probability, $D(t)$, defined as

$$D(t) = \frac{||P(t) - P_{eq}||}{||P(0) - P_{eq}||}. $$

(18)

Here, the 1-norm of a vector $P$ is defined as

$$||P|| = \sum_{a=1}^{K} |P_a|. $$

(19)

Then, the average of $D(t)$ over 100 trajectories, denoted as $\langle D(t) \rangle$, is taken. First, the results of different implementations but with the same move set MS$_3$ are shown in Fig. 3(a). The results indicate that while the implementation with $p_{ba}^{(0)} = 1$ and the acceptance ratio of Eq. (9) fail to converge towards thermal equilibrium, both Collet’s method and the new proposal show clearly the convergent tendency. In addition, the speed of convergence for Collet’s method is slower than that for the new scheme. The simulation results of the new method with different move sets are also shown in Fig. 3(b), and they indicate that the move sets MS$_2$ and MS$_3$ converge more rapidly than MS$_1$.

IV. RELAXATION TIME AND MOVE SET

Since the kinematics of foldings strongly depend on temperature, we analyze the relaxation time towards thermal equilibrium at various temperatures to clarify further the characters of different move sets. In general, the time scale of relaxation can be obtained by calculating the eigenvalues of the matrix $M$ of Eq. (6). The eigenvector corresponding to the zero eigenvalue, i.e., to the infinite relaxation time, is the equilibrium probability distribution $P_{eq}$ and the longest finite relaxation time $t_L$ can be found from the largest non-zero eigenvalue. In this section, based on the master equation of Eq. (5) and the inherent structure of conformation network, we give the upper bounds of the relaxation times, which is larger than $t_L$, for different move sets at various temperatures.

The basic idea in calculating the upper bound of relaxation time is as follows. Upon taking a node, say, $\mu$, as the referential target, there exists different paths for a node $a$ connecting to $\mu$. Subsequently, the transition rate associated with a path can be expressed as the product of the rates
between all pairs of two adjacent nodes along the path. First, we estimate the transition rate from \( \alpha \) to \( \mu \) by summing all the rates of the paths with different edge numbers up to \( l_{\text{max}} \). Note that \( l_{\text{max}} \) is the largest value of the minimal edge numbers connecting nodes to \( \mu \). Then, the minimum of the rates from all nodes to \( \mu \) is taken as the relaxation rate of the system. Moreover, the effect of possible loop paths is also considered. The results are essentially the consequence of the theorem shown in the Appendix. Based on the theorem, the 1-norm deviation \( D(t) \) of Eq. (18) satisfies the relation

\[
D(t) < \exp\left( -\frac{t}{t_c} \right),
\]

where

\[
t_c = \frac{1}{f_\mu} \exp(d_{\text{max}}),
\]

\[
d_{\text{max}} = \max \left\{ d'_a = \sum_{\beta \neq \alpha, \beta \neq \alpha} w_{\beta \alpha}, 1 \leq \alpha \leq K \right\},
\]

\[
f_\mu = \min \left\{ f'_a = \left( \sum_{k=0}^{l_{\text{max}}} \frac{W^k}{k!} \right)_{(\mu, \alpha)}, 1 \leq \alpha \leq K \right\},
\]

\[
l_{\text{max}} = \max \{ l_{\alpha}, 1 \leq \alpha \leq K \},
\]

where \( W \) is the off-diagonal part of the matrix \( M \) of Eq. (6), and \( l_{\alpha} \) is the minimal edge number connecting \( \alpha \) to \( \mu \). Note that the approximation \( \ln(1 + x) \approx x \) for small \( x \) has been used in obtaining Eq. (21). In general, the \( f_\mu \) value of Eq. (23) may depend on the referential target \( \mu \). However, conformation networks are quite close to random networks\(^8\) for which, the \( f_\mu \) values with respect to different \( \mu \) are expected to be indistinguishable. The relaxation time \( t_R \) can be defined as

\[
D(t) = \exp\left( -\frac{t}{t_R} \right).
\]

Then, the \( t_R \) value should be larger than the longest finite relaxation time \( t_L \), and the \( t_c \) of Eq. (20) is an upper bound of \( t_R \), namely, \( t_c > t_R > t_L \). Thus, the \( t_c \) value may serve as an alternative measure signifying the time scale of convergence toward thermal equilibrium.

We take the native state as the referential target and use the \( w_{\alpha \beta} \)'s obtained from our method for the calculations shown in the following. First, we choose the sequence of 12 monomers, \( \text{PHPPP} \), to compare the values of \( t_c \) and \( t_L \). The sequence has 15,037 conformations with the native state shown in the inset of Fig. 4(a). The \( t_c \) values are obtained by first symmetrizing the matrix \( M \) of Eq. (6) and then diagonalizing the symmetric matrix to obtain the eigenvalues. On the other hand, the \( t_c \) values are calculated from Eq. (21) with \( l_{\text{max}} = 24 \) (MS1), 10 (MS2), and 8 (MS3). The results are plotted as \( t_c \) versus \( T \) and \( t_c \) versus \( T \) for different move sets in Figs. 4(a) and 4(b), respectively. The \( t_c \) values for the sequence of 16 monomers given in Sec. II are also calculated with \( l_{\text{max}} = 49 \) (MS1), 14 (MS2), and 12 (MS3), and the plot of \( t_c \) versus \( T \) is given in Fig. 5. The results of the two sequences show that the values of \( t_c \) as well as \( t_R \), for the two move sets MS2 and MS3 are about the same, but the values for MS1 is drastically larger than those for MS2 and MS3. Note that the \( t_c \) values for MS2 is slightly larger than those for MS1 for the sequence of 12 monomers, and the

![FIG. 4. (a) The longest finite relaxation time \( t_c \) vs temperature \( T \) and (b) the upper bound of relaxation time \( t_R \) vs temperature \( T \) for the sequence of 12 monomers, \( \text{PPHPHPPPHPHPH} \), with the move sets MS1 (circles), MS2 (triangles), and MS3 (black dots). The inset of Fig. 4(a) shows the native conformation of the sequence.](image)

![FIG. 5. The upper bound of relaxation time \( t_R \) vs temperature \( T \) for the “proteinlike” sequence of 16 monomers specified as \( \text{PHPPP} \), with the move sets MS1 (circles), MS2 (triangles), and MS3 (black dots).](image)
situation is reversed for the sequence of 16 monomers. From this, we may expect that the relaxation time for MS2 is less than that for MS3 for long chains. Thus, the large-scale rigid rotations play a dominant role in bringing the system to thermal equilibrium. Moreover, the results of Fig. 4 indicate that the upper bound $t_r$ is many orders of magnitude larger than the $t_L$ value.

V. RIGIDITY-FACTOR AND KINETIC TRAP

The kinetic traps, referring to the states located at only few steps away from the native state but strongly prohibiting the folding into it, are important for understanding the folding kinetics. Their identifications usually rely on the distinction of local free energy minima. Here, by employing the algorithm of continuous time Monte Carlo simulations to the frame of conformation networks, we introduce the rigidity factor $\rho$ associated with a node to distinguish obstacle states. The $\rho$ value of a node is proportional to the average time duration of a state. Thus, a conformation with larger $\rho$ value is more difficult to deform to other conformations. The master equation, Eq. (4), indicates that the probability of the transition from $\alpha$ to any of the other states occurring in the time interval between $\tau$ and $\tau+d\tau$ is $(\Sigma_{\beta} w_{\alpha\beta})d\tau$, while in contrast the corresponding probability of nonoccurrence is $1-(\Sigma_{\beta} w_{\alpha\beta})d\tau$. Denoting further the probability of the nonoccurrence of the transition from $\alpha$ to other states for $t_r+r$ with $Q_{\alpha}(r)$, we have

$$\ln[Q_{\alpha}(\tau)] - \ln[Q_{\alpha}(\tau - d\tau)] = \ln[1 - w_{\alpha}^S d\tau],$$

which implies

$$\frac{d \ln[Q_{\alpha}(\tau)]}{d\tau} = -w_{\alpha}^S,$$

with $w_{\alpha}^S = (\Sigma_{\beta} w_{\alpha\beta})$. This yields

$$Q_{\alpha}(\tau) = \exp(-w_{\alpha}^S \tau),$$

where the initial condition, $Q_{\alpha}(0)=1$, is used. Then, the quantity $\bar{P}_{\alpha}(\tau)d\tau$, defined as

$$\bar{P}_{\alpha}(\tau)d\tau = Q_{\alpha}(\tau)w_{\alpha}^S d\tau,$$

is the probability that the transition from $\alpha$ to other states does not occur for $t_r+r$ but occurs in the time interval between $\tau$ and $\tau+d\tau$. Thus, the mean value of the time interval between reaching and leaving the node $\alpha$ is

$$\langle \tau_{\alpha} \rangle = \int_0^{\infty} \tau \bar{P}_{\alpha}(\tau)d\tau,$$

which yields $\langle \tau_{\alpha} \rangle = 1/w_{\alpha}^S$. For comparison, it is instructive to introduce the rigidity factor $\rho_{\alpha}$ associated with a node $\alpha$, defined as the ratio of $\langle \tau_{\alpha} \rangle$ to its average over all nodes,

$$\rho_{\alpha} = \langle \tau_{\alpha} \rangle \left( \frac{1}{K} \sum_{\alpha=1}^{K} \langle \tau_{\alpha} \rangle \right)^{-1}.$$

Consequently, a node with $\rho > 1$ in general possesses a rigidity factor greater than the average and requires more time to deform to other conformations.

The employment of rigidity factor can be shown useful for the coordination of the obstacle states during a folding process. To achieve this, the minimal edge number $l_{\alpha}$ from a node $\alpha$ to the native state fulfills the specification of metric factor and unambiguously defines the conformational distance. While the correspondences of nodes to the conformational distances are not unique, it is more appropriate to deal with the averaged rigidity factor equipped with a definite distance, say, $\langle \rho \rangle$. In calculating the $\rho$ values of Eq. (31), we use the $w_{\alpha\beta}$’s given by Eqs. (8), (16), and (17). Moreover, the characterization of degree of disorder at distance $l$ can be furnished by the calculation of entropy

$$S_l = -\sum_{\alpha} P_{eq,\alpha} \ln P_{eq,\alpha},$$

where the sum is taken over all nodes $\alpha$ with the same $l$.

The results of $\langle \rho \rangle$ versus $l$ and $S_l$ versus $l$ are shown in Fig. 6 for the model sequence of 16 monomers with different move sets at the folding temperature $T_f=0.5$. There are some important features of folding revealed from the results. First, as shown in Fig. 6 the entropic bottleneck of folding appears at $l=5$ (MS1) and 3 (MS2 and MS3). Then, we can identify the kinetic trap as the one with the largest rigidity factor $\rho$ among all nodes associated with the corresponding $l$ of the bottleneck. For the model sequence, the kinetic traps for different move sets happen to be the same, but this needs not to be generally true when other sequences are in consideration. As depicted in Fig. 7, the pathways from the trap to the native state remain unique and indistinguishable for MS2 and MS3; however, in contrast, there are two of such paths found diastatic for MS1. In addition, the rigidity factors for most of the nodes in the path are less than 1. This implies that after crossing over the kinetic trap the folding is very fast. Figure 7 reveals that large-scale rigid rotations are very effective in searching the negative state, but the local moves may be more effective for other sequences. Though the results are obtained at $T_f=0.5$, the qualitative features remain to be the same for other temperatures.

VI. SUMMARY

Based on the master equation approach endowed with the geometric structure of conformation network, we present new methods for the study of the folding kinetics of lattice polymers. In particular, the characteristic features of different move sets used in lattice Monte Carlo simulations are analyzed throughout the study. Firstly, we give the geometric structures of conformation networks for different move sets from which we propose a new method of implementing the Metropolis algorithm. The method is shown to satisfy the condition of detailed balance and to bring the distribution to thermal equilibrium very efficiently. Then, we use the master equation to show a theorem which can be used to place an upper bound of relaxation time. The upper bound is many orders of magnitude larger than the longest finite relaxation time obtained from the largest nonzero eigenvalue of the transform matrix of the master equation. Finally, we combine the new implementation with Gillespie algorithm of continuous time Monte Carlo simulations to introduce a more effec-
tive quantity, called rigidity factor, to identify the kinetic traps. Then, the folding kinetics can be exhibited clearly by assigning the rigidity factors to conformation distances. Here, the conformation distance from the native state is defined as the minimal edge number in the corresponding conformation network.

The new methods presented in this paper can be applied to the kinetic study of lattice polymers to obtain useful insights.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Council of Republic of China (Taiwan) under Grant No. NSC 95-2212-M-033-005. The authors thank the National Center for High Performance Computing for providing the computing facilities.

APPENDIX: THE DERIVATION OF EQUATIONS (20)–(23)

For the proof of Eqs. (20)–(24) we first give a proposition and its proof as follows.

Proposition

Consider a $K \times K$ stochastic matrix $S$ with the elements satisfying the properties: all $s_{\alpha\beta} \geq 0$ and $\Sigma_{\alpha=1}^{K} s_{\alpha\beta} = 1$ for $\forall \beta$. By defining the $\Gamma$ value of a stochastic matrix $S$ as

$$\Gamma(S) = \sum_{\alpha=1}^{K} s_{\alpha\min,\beta} \text{ and } s_{\alpha\min,\beta} = \min\{s_{\alpha\beta}, 1 \leq \beta \leq K\}, \quad (A1)$$

we have

$$||S \cdot (P - P')|| \leq (1 - \Gamma(S))||P - P'||, \quad (A2)$$

where $P$ and $P'$ are solutions of Eq. (5).

Proof: First we notice that

$$||S \cdot (P - P')|| = \sum_{\alpha=1}^{K} \left| \sum_{\beta=1}^{K} s_{\alpha\beta}(P_{\beta} - P'_{\beta}) \right|. \quad (A3)$$

By using the normalization condition, $\Sigma_{\beta} P_{\beta} = \Sigma_{\beta} P'_{\beta} = 1$, Eq. (A3) can be rewritten as

$$||S \cdot (P - P')|| = \sum_{\alpha=1}^{K} \left| \sum_{\beta=1}^{K} (s_{\alpha\beta} - s_{\alpha\min})(P_{\beta} - P'_{\beta}) \right|. \quad (A4)$$

Based on Eq. (A1), we have $s_{\alpha\beta} - s_{\alpha\min} \geq 0$ for $\forall \beta$, and thence
By interchanging the two sums on the right-hand side of Eq. (5) and by using the property \( \sum_{a=1}^{K} s_{a\beta} = 1 \), we obtain the inequality of Eq. (2). This completes the proof of the proposition.

The proposition is used to prove the result of Eqs. (20)–(23), which is expressed as the theorem below.

**Theorem**

For the probability distribution \( \mathbf{P}(t) \) evolved from initial distribution \( \mathbf{P}(0) \) via Eq. (5), there exists a \( \lambda \) with \( 0 \leq \lambda < 1 \) such that

\[
\| S \cdot (\mathbf{P} - \mathbf{P'}) \| \leq \sum_{a=1}^{K} \sum_{\beta=1}^{K} (s_{a\beta} - s_{\min,a}) |P_{\beta} - P'_{\beta}|.
\]  
(A5)

By interchanging the two sums on the right-hand side of Eq. (A5) and by using the property \( \sum_{a=1}^{K} s_{a\beta} = 1 \), we obtain the inequality of Eq. (A2). This completes the proof of the proposition.

The proposition is used to prove the result of Eqs. (20)–(23), which is expressed as the theorem below.

**Theorem**

For the probability distribution \( \mathbf{P}(t) \) evolved from initial distribution \( \mathbf{P}(0) \) via Eq. (5), there exists a \( \lambda \) with \( 0 \leq \lambda < 1 \) such that

\[
\| \mathbf{P}(t) - \mathbf{P}_{eq} \| \leq \lambda^{[t]} \| \mathbf{P}(0) - \mathbf{P}_{eq} \|.
\]  
(A6)

for a totally connected conformation network, where \( \mathbf{P}_{eq} \) is the equilibrium probability distribution and \( [t] \) is the integral part of the evolution time \( t \).

**Proof:** The matrix \( M \) of Eq. (6) can be expressed as

\[
M = D + W,
\]  
(A7)

where \( D \) is the diagonal part with elements

\[
d_{a\beta} = -\delta_{a\beta} \left( \sum_{\gamma \neq a} w_{\gamma a} \right),
\]  
(A8)

and \( W \) is off diagonal with elements \( w_{a\beta} \). By employing \( d_{\max} \) of Eq. (22), we write...
\[ M = [\bar{D} + W] - d_{\text{max}} I, \]  
\[ \text{where } I \text{ is the identity matrix, and the diagonal matrix, } \bar{D} = D + d_{\text{max}} I, \text{ is semipositive. Using the form of Eq. (A9) for } M, \text{ we can express the stochastic matrix } \exp(M) \text{ as} \]
\[ \exp(M) = \exp(-d_{\text{max}} \sum_{k=0}^{\infty} \frac{1}{k!}(\bar{D} + W)^k), \]  
\[ \text{which implies the inequality} \]
\[ \exp(M) \geq \exp(-d_{\text{max}} \sum_{k=0}^{\infty} \frac{W^k}{k!}), \]
\[ \text{because } \bar{D} \text{ is semipositive. We notice that there always exists a path connecting an arbitrary node to the referential target, say, the node } \mu, \text{ for a totally connected network. By referring } l_{\text{max}} \text{ as the largest value of the minimal edge numbers connecting nodes to } \mu, \text{ we have} \]
\[ \left( \sum_{k=0}^{\infty} \frac{W^k}{k!} \right)_{\langle \mu, \alpha \rangle} \geq \left( \sum_{k=0}^{l_{\text{max}}} \frac{W^k}{k!} \right)_{\langle \mu, \alpha \rangle}. \]  
\[ \text{Then, followed from Eq. (A11) with the inequality of Eq. (A12) and based on the definition given by Eq. (A1) for the } \Gamma \text{ value of a stochastic matrix, we have} \]
\[ \Gamma[\exp(M)] > f_\mu \exp(-d_{\text{max}}), \]
\[ \text{with } f_\mu \text{ given by Eq. (23). This leads to } \Gamma[\exp(M)] > 0, \text{ because } W \text{ is semipositive. Using the proposition and the inequality of Eq. (A13), we then have} \]
\[ \|\exp(M) \cdot (P - P')\| < \lambda\|P - P'\|, \]
\[ \text{with} \]
\[ \lambda = 1 - f_\mu \exp(-d_{\text{max}}) \]
\[ \text{for two solutions of Eq. (5), } P \text{ and } P'. \text{ Because of } 0 \leq 1, \text{ we may use Eq. (A14) to define a map of contraction } F, \]
\[ F(P - P') = \exp(M) \cdot (P - P'), \]
\[ \text{which contracts the 1-norm difference between } P \text{ and } P'. \text{ By successively applying } n \text{ times the map } F, \text{ we obtain} \]
\[ \|\exp(nM) \cdot (P - P')\| < \lambda^n\|P - P'\|, \]
\[ \text{for } n = 1, 2, 3, \ldots \text{ By taking } P' \text{ of Eq. (A17) to be the equilibrium distribution } P_{\text{eq}} \text{ which satisfies} \]
\[ \exp(M) \cdot P_{\text{eq}} = P_{\text{eq}}, \]
\[ \text{we have} \]
\[ \|\exp(nM) \cdot P - P_{\text{eq}}\| < \lambda^n\|P - P_{\text{eq}}\|. \]
\[ \text{For time evolution from 0 to } t, \text{ the master equation, Eq. (5), gives} \]
\[ P(t) = \exp(tM) \cdot P(0). \]  
We express } t \text{ as the sum of integer } \lfloor t \rfloor \text{ and fraction } r, \]
\[ t = \lfloor t \rfloor + r. \]
\[ \text{Then, the inequality of Eq. (A19) gives} \]
\[ \|P(t) - P_{\text{eq}}\| < \lambda^{\lfloor t \rfloor}\|P(0) - P_{\text{eq}}\|. \]
\[ \text{Since } \exp(rM) \text{ is a stochastic matrix, we apply the proposition to obtain} \]
\[ \|\exp(\tau M) \cdot P(0) - P_{\text{eq}}\| < [1 - \Gamma(\exp(rM))]\|P(0) - P_{\text{eq}}\|. \]
\[ \text{By substituting Eq. (A23) into Eq. (A22) together with the fact } 1 - \Gamma(\exp(rM)) \approx 1, \text{ we obtain} \]
\[ \|P(t) - P_{\text{eq}}\| < \lambda^{\lfloor t \rfloor}\|P(0) - P_{\text{eq}}\|. \]
\[ \text{This completes the proof of the theorem.} \]