Conformation-networks of two-dimensional lattice homopolymers

Yu-Pin Luoa, Hung-Yeh Linb, Ming-Chang Huang a,*, Tsong-Ming Liaw c

a Department of Physics, Chung-Yuan Christian University, Chungli 320, Taiwan
b Institute of Electrophysics, National Chiao-Tung University, Hsinchu 300, Taiwan
c Computing Center, Academia Sinica, Nankang 115, Taiwan

Received 6 July 2005; received in revised form 6 June 2006; accepted 8 June 2006
Available online 19 June 2006
Communicated by C.R. Doering

Abstract

The effect of different Monte Carlo move sets on the folding kinetics of lattice polymer chains is studied from the geometry of the conformation-network. The networks have the characteristics of small-world: the local connections are more clustered than that of the corresponding random lattices, and the characteristic path lengths increase logarithmically with the number of nodes. One of the elementary moves, rigid rotation, has drastic effect on the geometric properties of the network. The move increases greatly the connections and reduces significantly the shortest path lengths between conformations. Including rigid rotation to the move set results in the increase of the dimensionality of the conformation space to the value about 4.

© 2006 Elsevier B.V. All rights reserved.

Protein folding is a complex process for which, a sequence of amino acids folds into a unique and stable structure in a relatively short time [1]. The lattice models have been used widely as coarse-grained models for the theoretical study of folding process [2–7]. In the lattice models, protein is viewed as a chain of $m$ monomers, and the conformations are given by all possible self-avoiding walks of the chain on a two or three-dimensional lattice. The energy of a conformation, in general, depends on the number of intrachain contacts, and how to assign the contact energy is model dependent. The kinetics of folding process can then be studied by Monte Carlo simulations for which, a move set is designed for the change of conformations. In principle, different move sets, satisfying the requirement of ergodicity, should reach the same equilibrium canonical distribution after sufficiently long time simulations. However, different move sets may yield different perspectives of folding kinetics. Chan and Dill analyzed the folding kinetics of two different move sets for 2D homo- and heteropolymers by using the Metropolis transfer matrix method [5,6]. Their results indicate that a move set adopted for the study affects strongly the kinetic sequence of foldings and the shape of the energy landscape. Same conclusions were also given by Hoang and Cieplak [8] via the comparison between the dynamics of three different move sets. Thus, understanding the nature of a move set is essential for the interpretation of simulation results.

In this Letter, we explore the characteristics of different move sets via the analyses of the corresponding conformation-networks [9,10] for the 2D homopolymers with monomers $m \leq 16$. Though the chain lengths considered are relatively short, the networks can be constructed by exact enumeration. Scala, Amaral, and Barthélémy studied various networks obtained from the mappings of a particular conformation space, and showed that the geometric properties are similar to those of small-world networks [10]. This leads to the question, whether the conformation-networks obtained from different move sets all show the small-world characteristics. There appears two essential characteristics for a small-world network: (i) the local connection is more cliquey than that of random lattices, and (ii) the characteristic path length increases logarithmically with the number of nodes [11,12]. Thus, firstly we analyze the characteristic path lengths and clustering coefficients of the net-
works. To further differentiate the networks, we compute and compare the degree distributions of a node, the correlations between degrees of nearest-neighbors, and the distributions of the distances between two nodes. Finally, we also discuss the stability of the networks.

For the dynamical simulations of lattice polymers, the typical elementary moves include the end flip (ef), corner shift (cs), crankshaft (cr) and the rigid rotation (rr), as shown in Fig. 1. Here, some specially designed moves, such as snake move [8], are excluded from the consideration. We focus the study on the move sets $S_1$, $S_2$, and $S_3$, defined as follows. The conventional move set $S_1$ consists of ef, cs and cr [7,13,14], based on locality. However, the ergodicity cannot be generally satisfied for $S_1$ [5,6,8]. 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 [15]. While ef itself can be viewed as short-scale rigid rotation, an ergodic move set, say $S_2$, can be achieved by simply combining ef with rr. Finally, the ergodic move set $S_3$ contains all the moves of four types.

For the construction of the network associated with a move set, firstly we identify all possible self-avoiding conformations of the chain of $m$ monomers as the nodes of the network associated with a move set. The node-number is denoted as $N_m$ for which, the degeneracy caused by the rotation and the mirror symmetry has been excluded. Two nodes are then connected by an edge if a move of the given move set can transfer one to the other. Thus, different move sets yield different edge distributions between the nodes and hence different networks. We refer the edge-number as $E_m$. The values of $N_m$ and $E_m$ for three networks with various numbers of monomers $m$ are listed in Table 1. Note that because all edges are undirected and have the same weight, the networks can be viewed as the folding networks in high temperature limit.

The edge-number associated with a node is also referred as the degree of the node, and the degree distribution $P(k)$ is defined as the probability for a node to have degree $k$. Then, the mean degree of a network is

$$\langle k \rangle = \sum_k k P(k),$$

which is equal to $2E_m/N_m$. The $\langle k \rangle$ values of different networks with various numbers of monomers $m$ are given in Table 1. The scaling of $\langle k \rangle$ with $N_m$ behaves as $\langle k \rangle = a + b \log(N_m)$ with $(a, b) = (3.79, 0.92)$ for $S_1$, $(3.07, 2.99)$ for $S_2$, and $(2.77, 4.01)$ for $S_3$, as shown in the insets of Fig. 2. Thus, the mean degree of the move set $S_2$ ($S_3$) is about two (two and half) times the value of $S_1$. A larger value of the mean degree of a network should give more throughway accessibility to the native conformation and reduce the chance of being trapped in local minimum in the folding process [5,6]. The results of $P(k)$ vs. $\Delta k = k - \langle k \rangle$ are shown in Fig. 2 for $S_1$, $S_2$, and $S_3$, respectively, with $m = 10, 12, 14$, and 16. Scala et. al. studied the sub-networks of $S_1$ for which, a sub-network is specified by a given end-to-end distance and generated by the moves, corner shift and crankshaft move [10]. Their results showed that the form of $P(k)$ is Gaussian. Then, we employ the Gaussian function,

$$P(k) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[ -\frac{(k - \langle k \rangle)^2}{2\sigma^2} \right],$$

to fit the data of $S_1$, $S_2$, and $S_3$, and the best fittings are given as the solid lines in Fig. 2. For $S_1$, the distribution agrees with the above Gaussian form in which, the variances of different $m$ are $\sigma = (0.5748)\sqrt{N_m}$. Comparing with the result of $S_1$, the distribution for $S_2$, shown in Fig. 2(b), does not fit so well, and the result of $S_3$, shown in Fig. 2(c), exhibits significant deviation, but obviously the distributions are not scale-free [16–18].

The deviation from the Gaussian form for the degree distributions reflects in the asymmetry between the distributions of low and high degrees, and the asymmetry can be clarified further by measuring the degree–degree correlations. The corre-
Fig. 2. The degree distribution, $P(k)$, versus $\Delta k = k - \langle k \rangle$ for the networks associated with different move sets: (a) $S_1$, (b) $S_2$, and (c) $S_3$. Here, $\langle k \rangle$ is the average edge number per node, and the solid lines are the best fittings of the Gaussian function given in the text. For each network, the plot of $\langle k \rangle$ vs. $\log(N_m)$ for the node number $N_m$ with the monomer number $m$ ranged from 8 to 16 is shown in the inset, and the straight solid line corresponds to the relation $\langle k \rangle = a + b \log(N_m)$ with the values of $a$ and $b$ given in the text.
where $k_i$ are listed in Table 1. The characteristic length of a network can be characterized by the average of the clustering coefficients of the nodes, denoted by $C_i$. The $C_i$ values for $S_1$, $S_2$, and $S_3$ with different $m$ values are listed in Table 1. The data shows $C_{S_1} > C_{S_2} > C_{S_3}$. For the network with the node-number $N$ and the average edge-number $k(k)$, the corresponding random network has the average clustering coefficient $C_{ran} \approx \langle k \rangle / N$. The results of the ratios $C / C_{ran}$ vs. the node-number $N_m$ for $S_1$, $S_2$, and $S_3$ are shown in Fig. 4 with logarithmic scales. Our results indicate that the average clustering coefficients of the conformation-networks are much larger than that of random networks. In particular, the network of $S_1$ is less random than that of $S_2$ and $S_3$. Thus, the kinematics based on $S_1$ have more chances to be trapped in some cliquy conformations than those based on $S_2$ and $S_3$.

We may define the minimum number of elementary moves required for transferring one node to the other as the distance between the two [5,6]. Thus, the distance $l$ between pairs of nodes is the minimum number of edges required to connect the two nodes. The distribution $P(l)$ gives the probability of distance $l$ between two randomly chosen nodes. The characteristic length of the network can be defined as the average of the distances of all node-pairs,

$$l = \sum l P(l).$$

The values of $\langle l \rangle$ for $S_1$, $S_2$, and $S_3$ with different $m$ values are listed in Table 1. The characteristic length of $S_2$ is about half of the length of $S_1$. For the distributions $P(l)$, the scaled plots of $P_{scaled}(l) = \sqrt{2\pi} \sigma P(l)$ vs. $\Delta l_{scaled} = (l - \langle l \rangle)/\sqrt{2} \sigma$ are shown in Fig. 5, where the solid lines are obtained by setting the variances $\sigma$ as $\sigma_{S_1} = 0.0489(m)^{1.7}$, $\sigma_{S_2} = 0.3057(m)^{0.6}$, and $\sigma_{S_3} = 0.5057(m)^{0.6}$, for $m$ monomers. The above variances are determined by first finding the least square fit to Eq. (2) to obtain $\sigma(m)$, and then taking the average over $\sigma(m)$ of different $m$. The distributions all agree with the Gaussian form of Eq. (2). The variance of $P(l)$ for $S_2$ is much smaller than that for $S_1$, and this implies that the distance between two nodes does not vary much for the networks based on $S_2$ and $S_3$.

For the small-world networks, there exists a cross-over size $N^* \approx \rho^{-1}$ such that the characteristic lengths $\langle l \rangle$ obey the finite-size scaling law [21–23],

$$\langle l \rangle = (N^*)^{1/d} f \left( \frac{N}{N^*} \right).$$

Fig. 3. The plot of $\bar{k}(k) - \bar{k}^*(k)$ vs. $k - \langle k \rangle$ for the networks $S_1$, $S_2$, and $S_3$ with $m = 16$. Here, $\bar{k}(k)$ is the average degree of the connected neighbors of a node with degree $k$, and $\bar{k}^*(k)$ is the result of $\bar{k}(k)$ for the absence of correlations.

The values of $\langle l \rangle$ for $S_1$, $S_2$, and $S_3$ to the average clustering coefficients of the corresponding random networks $C_{ran}$ versus $\log(N_m)$ with the node number $N_m$ and the monomer number $m$ ranged from 8 to 16.
Fig. 5. The scaled result of the distribution function of the shortest path lengths, \( P_{\text{scaled}}(l) = \sqrt{2\pi\sigma} P(l) \), versus \( \Delta_{\text{scaled}} = (l - \langle l \rangle) / \sqrt{2}\sigma \) for (a) \( S_1 \), (b) \( S_2 \), and (c) \( S_3 \) with \( m = 10, 12, 14, \) and 16. The averages of the shortest path lengths for all node-pairs, \( \langle l \rangle \), are given in Table 1, and the variances \( \sigma \) are given in the text. The solid lines are the results of the Gaussian form.

Fig. 6. The plots of the characteristic path length \( \langle l \rangle \) versus the logarithm of the node-number \( \log(N_m) \), for the networks associated with different move sets, (a) \( S_1 \), (b) \( S_2 \), and (c) \( S_3 \), where the monomer number \( m \) ranges from 5 to 16. The insets are the plots of \( \log(\langle l \rangle) \) versus \( \log(N_m) \) for the same data. The solid lines are the results of the limiting scaling forms given in the text.
due to the fact that the previous results are based on the networks with fixed end-to-end distances of the chain, such networks exclude the elementary move ef and are sub-networks of the networks of $S_1$.

Finally, we analyze the ability of attack and error tolerance of the network by studying the fragmentation caused by node-removal [24]. The nodes with higher degrees of connections are removed preferentially for the analysis of attack tolerance; and the nodes are removed randomly for the error tolerance. By removing a fraction $f$ of the nodes, we measure the fraction of nodes contained in the largest cluster, $S$, and the average node number, $\langle s \rangle$, contained in the fragmentary clusters excluding the largest one versus the fraction $f$ of the nodes removed for (a) attack and (b) error tolerance of the networks $S_1$, $S_2$, and $S_3$ with $m = 16$.

In summary, we divide the frequently used Monte Carlo moves into three different move sets, and construct the corresponding conformation-networks. The networks all have the characteristics of small-world: (i) the local neighborhood is more cliquish than that of random networks, and (ii) the characteristic path length increases logarithmically with the number of nodes. The dimensionalities of the conformation-spaces are $d \geq 3$. Our analyses also indicate that the networks are as robust as random graphs. Among different elementary moves, the rigid rotation has drastic effect on the geometric properties of the network: (i) it renders the connection distribution to be non-Gaussian, (ii) it reduces greatly the characteristic path length, and (iii) it drives the network more closer to random networks. Thus, the rigid rotation may change the folding kinetics significantly from that of the local moves, corner shift and crankshaft move.

Acknowledgements

This work was partially supported by the National Science Council of Republic of China (Taiwan) under the Grant No. NSC 93-2212-M-033-005. We thank the National Center for High-performance Computing and the Computing Centre of Academia Sinica for providing the computation facilities.

References

[22] A. Barrat, cond-mat/9903323.
    Erratum.