Study of Move Set in Cubic Lattice Model for Protein Folding

Jae-Min Shin* and Won Seok Oh

Molecular Design Laboratory, Hanhyo Institutes of Technology, 461-6 Chunmin-Dong, Yoosung-Ku, Taejun 305-390, Korea

Received: August 13, 1997; In Final Form: June 2, 1998

In cubic lattice model, we have devised a new move set that is fast and suitable for use in a protein folding study by Monte-Carlo simulations, where the number of moving monomers are more extended and flexible in pivot moves and crankshaft moves. In Monte Carlo simulations using 100 model proteins with 27-mers in a cubic lattice, this new move set shows much higher foldicity and faster mean first passage time than the most commonly used move set. From the results of lattice simulation at equilibrium conditions, it is found that the conformational space sampled by this new move set is almost the same as the one sampled with the commonly used move set. However, in the simulation at low temperature, the detailed kinetic folding paths of this new move set are shown to be different from that of the commonly used move set.

I. Introduction

Biological activity of a protein is associated with the native three-dimensional (3D) structure which is fully determined by its primary amino acid sequence.[1] The study of the protein folding mechanism, how the protein's native three-dimensional (3D) structure is formed in a given environment, has been one of the main research subjects.

A simulation study of protein folding requires a simplified protein model because of the complexity of the conformational states of real proteins. The lattice model is one of the simple computational models instructive in studying protein folding mechanisms.[2−4] As the lattice model greatly reduces the complexity of conformational space, an exhaustive conformational search becomes feasible.[5−7] Many detailed modeling studies on protein folding have been done with the lattice model, such as statistical thermodynamics,[2−4] prediction of 3D structures,[8,9] folding kinetics,[10−14] sequence design,[11,12,15,16] and cooperativity.[2−4,17−19] Although the simplified lattice model clearly loses the details of protein structures and functions,[20] it does not lose the physical essence of protein folding features,[2−4,21] such as the existence of cooperativity and the primary sequence pattern that determines its uniquely defined native 3D conformation.

In the lattice model, however, a computational strategy for conformational search is required in an attempt to find the global energy minimum state, which is commonly assumed to correspond to the native state of the protein.[2−4] The most common strategy in the lattice model used in the protein folding study is based on the Monte Carlo (MC) algorithm, in which a new conformation is accepted or rejected by the probability of Metropolis criteria,[22] $P = \min(1, \exp(-\Delta E/kT))$, where ΔE is the energy difference between the new and old conformation and kT is the Boltzmann constant multiplied by temperature. The conformation search by the Metropolis algorithm results in a Markov chain so that long time simulation produces a thermodynamic ensemble that can be used in the analysis of various useful equilibrium properties.

Another important aspect in the computational strategy is the "move set" of the lattice chain because the sampling itself in the MC simulation is performed by a move selected in the move set used. The widely used move sets[11,12] are usually based on three types of moves: one-bead terminal swing, one-bead corner flip, and two-bead crankshaft motions[23] (see Figure 1). A one step MC move in the lattice model corresponds to one randomly chosen move among these three predefined move types if this new conformation is topologically allowed and self-excluded in a given lattice. Based on these three move types, all the conformational states can be explored in a given lattice space. However, this move set may be inadequate to simulate the protein folding motion because the real protein folding is usually thought to depend not only on the local motion which can be simulated by these three move types but also on the nonlocal motion such as diffusion/collision motion.[24−26] Thus, folding simulations based on only local moves may produce a number of trivial back-and-forth trajectories during MC runs. Although there are many different versions for move sets,[11−13,27,30] one of the important motivations of our study is to incorporate proper nonlocal motions which have the property of tending to preserve local favorable conformations in the protein folding.[27] In the present study, we extend the previously used move set by considering nonlocal moves and we discuss the advantages of this move set by comparing to the commonly used move set.

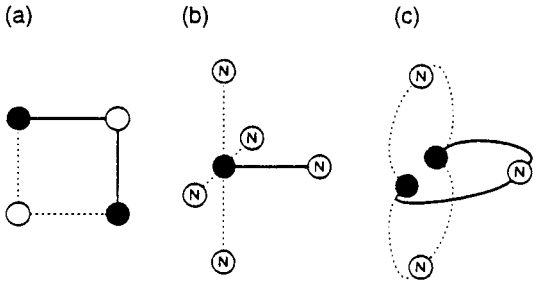

Figure 1. Schematic representation of the move set used in this study: (a) one-bead corner-flip move; (b) pivot move; (c) crankshaft move. Solid lines represent the current state, and dashed lines show the state after the moves. Filled circles are the beads that are not to move, and open circles are the beads to move. N in pivot and crankshaft moves represents the number of beads to move. In move set 1 (MS1), $N = 1$ for a pivot move and $N = 2$ for a crankshaft move. In move set 2 (MS2), N is a random number between 1 and $N_{\text{max\_pivot}}$ for a pivot move, and N is a random even number between 2 and $N_{\text{max\_crank}}$ for a crankshaft move (see text).

* Corresponding author. Current address: Lab. Mol. Biol., Bldg. 37, Room 4B15, NIH, Bethesda, MD 20892. E-mail: jms@rose.hci.nih.gov. Fax: (301) 402-1344.

S1089-5647(97)02648-5 CCC: $15.00
Published on Web 07/29/1998

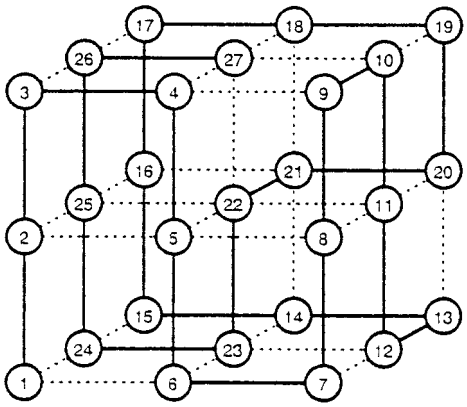

Figure 2. Native conformation (global energy minimum conformation) of sequence 12 (see Table 1) in a $3 \times 3 \times 3$ cubic lattice. There are 27 monomers linked through solid lines and 28 native contacts drawn by dashed lines.

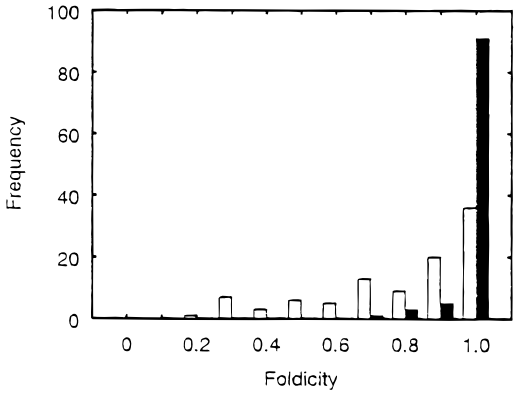

Figure 3. Distribution of foldicities of 100 model sequences for MS1 (open bar) and MS2 (filled bar) at $T = 0.45$. Foldicity is defined as the number of MC runs out of 10 runs that find the native state within 5×10^5 MC steps.[10,11]

II. Lattice Model

A. Move Set. As shown in Figure 1, widely used lattice models[2,3,11,12] usually use three types of local chain moves as a move set; one-bead terminal swing, one-bead corner flip, and two-bead crankshaft moves, respectively. In this study, we termed this move set a move set 1 (MS1). In our new move set called move set 2 (MS2), the number of moving beads in both pivot and crankshaft moves is more extended and can be varied within upper limits, $N_{\text{max\_pivot}}$ for a pivot move and $N_{\text{max\_crank}}$ for a crankshaft move, respectively. In a pivot move, for example, a monomer is randomly chosen in the range (1, $N_{\text{max\_pivot}}$), and this monomer is randomly assigned to be a pivot point for either amino- or carboxy-terminal moving segments of the lattice chain with an equal probability, and then these moving segments are rotated by $\pm 90°$ about a randomly chosen axis. The pivot move in MS2 is the same as a rigid rotation which was previously used in many lattice models by other authors.[13,27] In the following, we briefly describe the computational method used in MS2.

a. One-Bead Corner Flip Move. In a one-bead corner flip move, the new position of a randomly chosen monomer i, r_i, can be generated from the positions of three monomers, $i - 1$, i, and $i + 1$ by the equation $r_i^{n+1} = r_{i-1}^n + r_{i+1}^n - r_i^n$, where r_i^{n+1} is the position of monomer i at MC step $n + 1$.[23] Thus, when three monomers $i - 1$, i, and $i + 1$ are not in the same axis, the new position of the ith monomer flips to the diagonally opposite corner site.

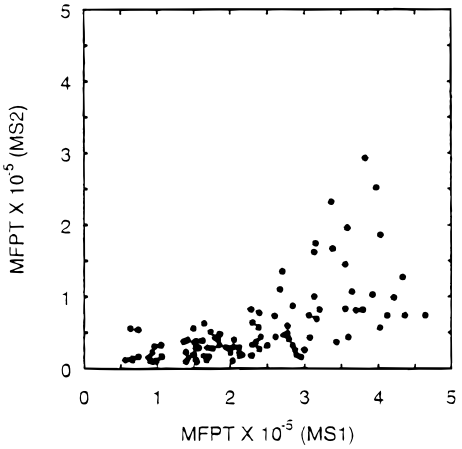

Figure 4. Comparison of MFPTs[13,15] between MS1 and MS2 for 100 model sequences. In the calculation of MFPT, the cases that do not find the native state within 5×10^5 MC steps are considered as if the native states were found at 5×10^5 MC steps.

b. Pivot Move. A pivot move about a randomly chosen monomer p is obtained by using the direction cosine method.[31] For example, the direction cosine for rotation about the z axis by 90° has the simple form

$$R_z = \begin{bmatrix} \cos(90) & \sin(90) & 0 \\ -\sin(90) & \cos(90) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus, pivot motion at the monomer p about z axis by 90° can be obtained by the operation $r_i^{n+1} = R_z(r_i^n - r_p^n) + r_p^n$, where index i is applied to all the moving beads. Pivot rotation about other axis can be obtained in a similar way using the direction cosines.

c. Crankshaft Move. For the crankshaft move, we randomly choose a monomer i in the range between (1, $N_{\text{chain}} - 3$), where N_{chain} is the total number of monomers in a lattice chain, 27 for our case, and then we choose a monomer j that is in contact with i and is in closest position in sequence from i by running index j from $i + 3$ to min($i + N_{\text{max\_crank}}$, N_{chain}). Rigid body rotation using the direction cosine about the common axis of these two selected monomers is applied to the range of monomers from i to j by $\pm 90°$. Thus, the crankshaft move in MS2 is defined by $\pm 90°$ rotation of the smallest "loop" starting from a selected monomer i, where the "loop" means the range of monomers where the two ends are in contact.

d. Selection of Move Type. For the move type selection in MC simulations, we examined various combinations of trial probabilities for each of the three move types. After we carried out a number of pilot experiments with MS2, we assigned the probability of 0.4, 0.2, and 0.4 for one-bead corner-flip, pivot, and crankshaft moves, respectively. For MS1, we used the same selection probability in ref 11 for the three local moves. Although different assignment of the selection probabilities for the move types may affect the detailed dynamics of the MC folding simulation, the results of this study would not be affected very much. The parameters $N_{\text{max\_pivot}} = 13$ and $N_{\text{max\_crank}} = 12$ were used with MS2.

B. Selection of Model Sequences and Native Conformations. As shown in Figure 2, a simple cubic lattice chain with 27 monomers is used as a model protein. Following the methods of ref 15, the sequences of model proteins are designed to be in a global energy minimum state when its conformation is one of the 103 346 maximally compact conformations[6,7,14] in a $3 \times 3 \times 3$ cubic lattice, and the global energy minimum

TABLE 1: Examples of Model Protein Sequences and Corresponding Native Energies Used in This Study

no.	sequence																										E_{nat}
1	E	V	W	R	A	A	W	K	P	M	W	Y	Y	R	V	K	E	G	S	S	G	G	I	M	N	K	G	−20.72
2	A	I	Q	M	G	Y	T	K	Y	S	E	I	C	K	E	G	Y	C	W	M	G	G	W	R	L	M	L	−21.28
3	R	S	G	I	S	T	A	E	K	P	S	G	W	E	C	M	G	P	L	M	C	I	R	C	W	P	L	−21.44
4	W	R	G	C	G	A	W	G	V	G	Q	G	C	K	C	L	E	M	D	R	W	M	Q	V	W	M	I	−21.91
5	R	G	K	M	C	P	C	L	W	E	R	M	W	E	R	M	A	A	W	E	G	C	A	C	I	A	I	−24.28
⋮																												
12	E	G	A	W	C	E	K	C	M	P	C	E	A	M	E	M	L	M	Q	A	W	C	K	K	G	I	M	−23.07
⋮																												
32	M	Q	A	D	K	M	G	R	K	P	I	E	W	K	W	G	E	G	M	K	E	L	Q	W	P	G	G	−23.41
⋮																												
96	C	E	C	W	K	E	M	I	M	F	G	C	G	G	K	P	L	G	A	T	T	E	Q	I	L	P	K	−23.35
97	T	L	I	Q	T	E	K	L	Y	E	K	M	G	E	I	Y	W	A	I	G	G	G	K	G	W	E	K	−23.67
98	E	Q	E	L	C	A	W	C	P	I	T	K	L	Q	C	H	I	K	E	H	G	M	A	G	P	S	I	−21.60
99	G	C	C	N	G	N	M	R	Y	G	D	G	P	A	C	S	A	R	D	W	M	Y	A	C	N	W	N	−21.27
100	C	M	R	I	E	I	L	C	F	Q	E	M	L	K	E	S	R	K	A	E	W	G	P	C	W	R	G	−23.35

TABLE 2: Foldicity and Mean First Passage Time (MFPT) for MS1 and MS2[a]

	foldicity		MFPT × 10^{-3}	
T	MS1	MS2	MS1	MS2
0.35	4.80	9.58	357	120
0.40	7.07	9.77	268	73
0.45	7.98	9.86	234	59
0.50	7.34	9.60	261	88
0.55	5.52	8.81	326	147

[a] Here, foldicity is defined as the number of successful MC runs that find native conformation within 5×10^5 steps among 10 independent MC runs for each sequence.[10,11] The values for foldicity and MFPT[13,15] are averaged from 100 model sequences. In the calculation of MFPT, the runs that do not find the native states within 5×10^5 MC steps are considered as if the native states were found at 5×10^5 MC steps (see text).

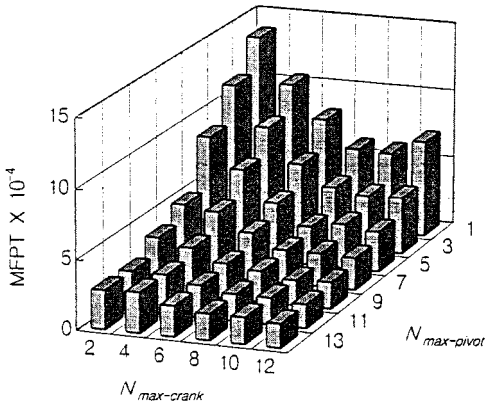

Figure 5. Dependence of MFPTs on the two parameters for move sizes. The MFPT of MS1 corresponds to the bin where $N_{max\_pivot} = 1$ and $N_{max\_crank} = 2$, while the MS2 in this study is the bin where $N_{max\_pivot} = 13$ and $N_{max\_crank} = 12$.

conformation is regarded as the native conformation of this sequence.[2−4] The potential energy of a conformation is calculated by eq 1:

$$E = \sum_{i<j} B_{ij}\Delta(r_i - r_j) \tag{1}$$

where E is total energy, B_{ij} is the interaction energy parameter between monomers i and j, and $\Delta(r_i - r_j)$ is 1 when monomer i and j are in contact and is 0 otherwise. The interaction energy parameter B_{ij} is defined in eq 2:

$$B_{ij} = \epsilon_{ij} + B_0 \tag{2}$$

where ϵ_{ij} is Miyazawa−Jernigan parameter.[32] B_0 is used as a parameter for background potential, which has the meaning of overall attraction, thus forcing the lattice chain to be in a more compact conformation.[15] Following ref 15, we used −0.32 for B_0.

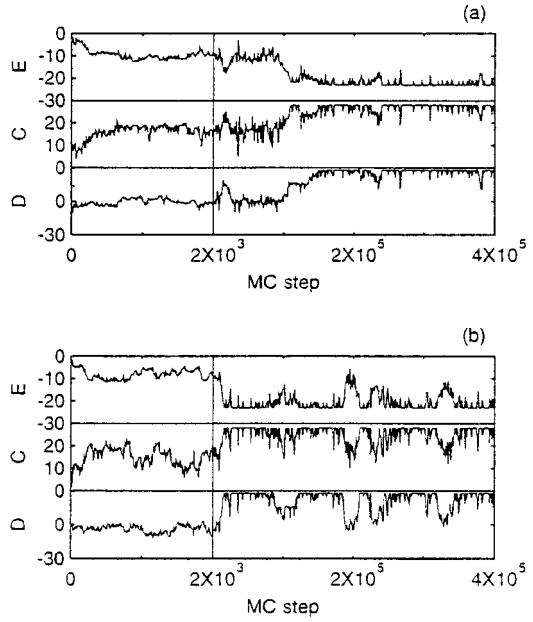

Figure 6. Typical MC simulation trajectories of MS1 and MS2 at $T = 0.45$. E is energy, and C is the total number of contacts. D is the parameter defined in the text and is used for structural monitoring; that is, the native state corresponds to the state at $D = 28$. For convenience, different scales are used in MC steps.

For the design of 100 model protein sequences, we randomly choose 100 conformations among 103 346 maximally compact conformations of a 27-mer cubic lattice chain.[6,7] Then, the sequences are designed by minimizing the relative value of the native energy, E_{rel}, in eq 3, as was proposed by Shakhnovich et al.[15,33]

$$E_{rel} = \frac{E_{nat} - E_{ave}}{\sigma} \tag{3}$$

where E_{nat} is the energy of the native conformation and E_{ave} is the average energy of all the possible compact conformations of a given sequence, i.e., $E_{ave} = 28e_{ave}$, where e_{ave} is the average energy per one topologically possible contact. Standard dispersion σ is defined by eq 4:

$$\sigma = [(\sum_{i,j}(B_{ij} - e_{ave})^2)/\mathrm{N}]^{1/2} \tag{4}$$

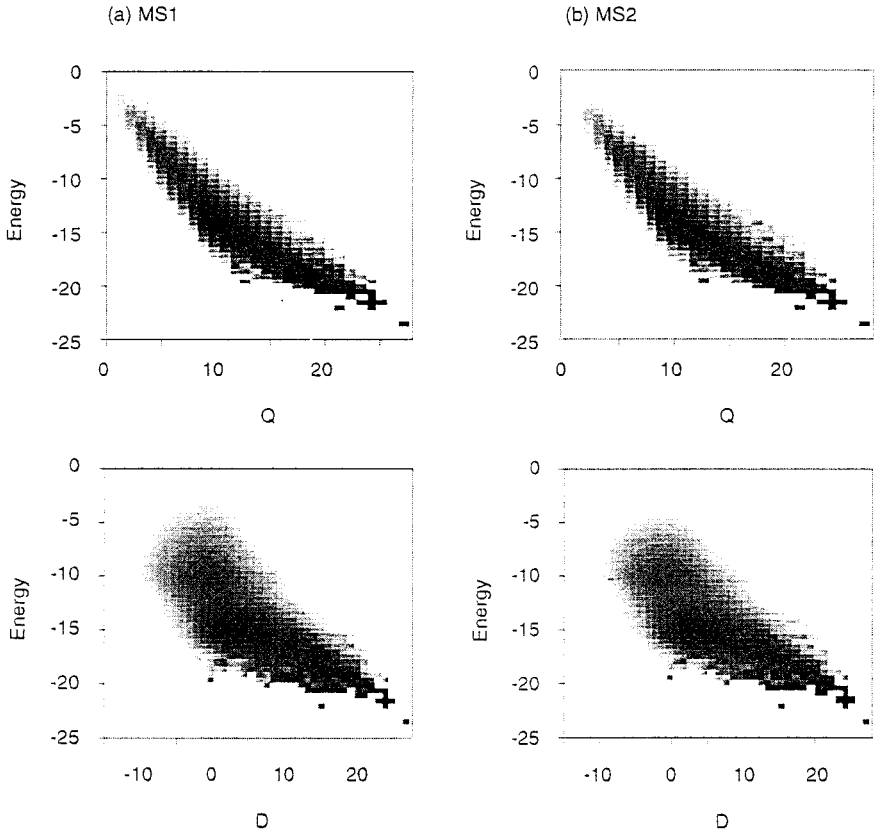

Figure 7. Density of states for sequence 12 at $T = 0.45$; (a) MS1; (b) MS2. Each bin represents the log scale of the density of states that spans 0.5 energy unit and one unit in reaction coordinates, Q or D. The average is made from more than 1000 independent MC runs.

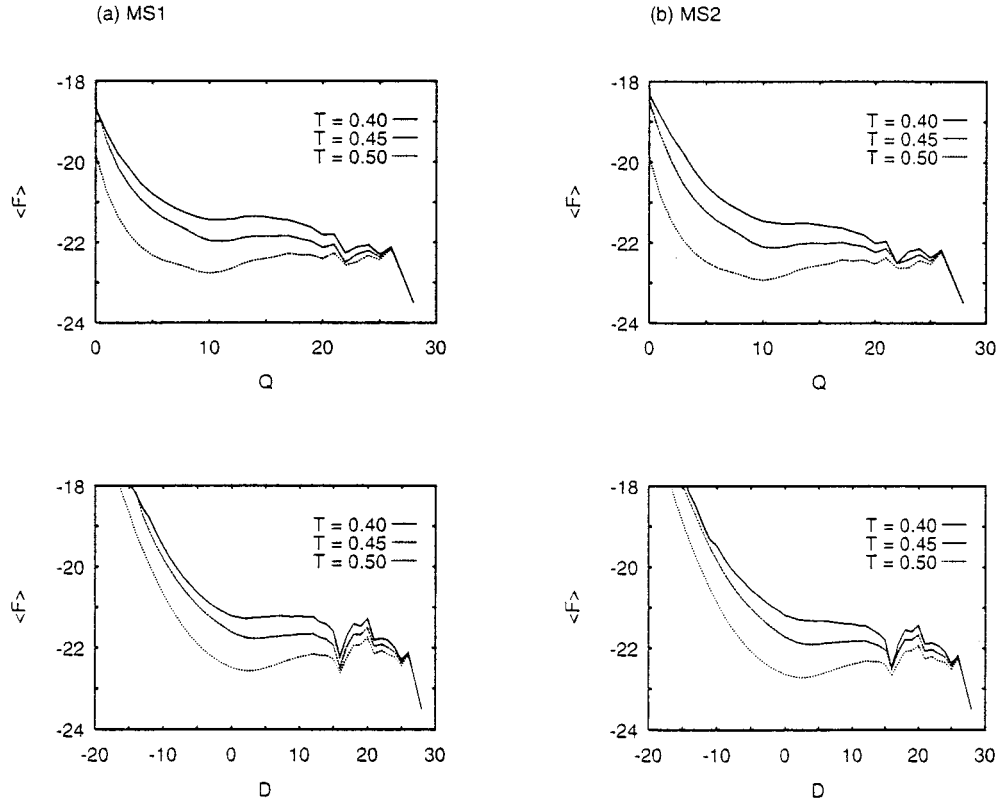

Figure 8. Free energy profiles for sequence 12 at three different temperatures, $T = 0.4$, $T = 0.45$, and $T = 0.5$, respectively. Free energy is calculated from the analysis of the average density of states shown in Figure 7 by the methods described elsewhere.[10,14]

where indices i and j run over all the possible contacts and N is their number. In a given sequence, minimization of E_{rel} is driven by a random mutation of one residue and continues until 100 consecutive mutations do not find lower energy. We confirm that there are no incorrect conformations with lower energy than the target native conformation by performing 10

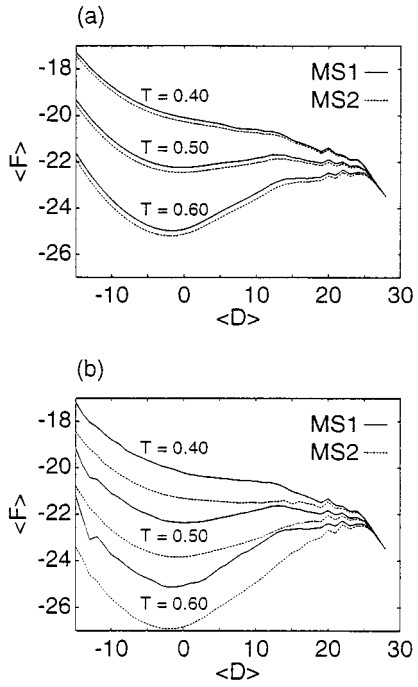

Figure 9. Folding profiles of MS1 and MS2 for sequence 32 which are obtained from the simulation at $T = T_f = 0.49$ (a) and at $T = 0.40$ (b). A statistical mechanical relation[36] is applied to show the profiles at three different temperatures. For sufficient equilibration at these temperatures, a number of long MC simulations (20 independent runs with more than 10^8 MC steps) are carried out for both MS1 and MS2. For this sequence, MS1 shows slightly higher T_f than MS2, i.e., T_f(MS1) = 0.50 and T_f(MS2) = 0.49. T_f is defined by the relationship $\exp(-E_N/T_f)/Z = 0.5$, where E_N is the native energy and Z is the partition function obtained from the simulation.

independent 1×10^7 steps of MC simulations for each sequence. Table 1 shows some examples of model sequences and corresponding native energy values used in this study.

III. Results and Discussion

A. Foldicity and Mean First Passage Time of MS1 and MS2. For all 100 model sequences, 10 independent MC simulations are carried out at five different temperatures, from $T = 0.35$ to $T = 0.55$ with an interval of 0.5. Foldicities and mean first passage times (MFPTs) of MS1 and MS2 are analyzed for all 100 model sequences and are summarized in Table 2. In this study, foldicity[10,11] is defined as the ratio of the number of successful MC runs that found the native conformation within 5×10^5 MC steps to the number of total MC runs tried. The MFPT[13,15] is the average MC folding steps at which the native conformation is first found during the MC simulation starting from an initial random conformation. As shown in Table 2, MC simulations at $T = 0.45$ result in the best foldicities and the fastest MFPTs in both MS1 and MS2, and also it can be seen that MS2 shows considerably better performance than MS1.

Figure 3 shows the distribution of foldicities at $T = 0.45$. This figure indicates that the foldicities of all the model sequences are greater than 0.7 in MS2, while in MS1 the foldicities of the same model sequences are widely distributed. Furthermore, for MS1, there are more than 10 sequences that show the foldicities of less than 0.5, meaning more than 5 MC runs out of 10 runs failed in finding the native structure with 5×10^5 MC steps. In MS2, 95 sequences out of 100 sequences were found to show the foldicity of 1.0, indicating that, for those sequences, MS2 is likely to guarantee that every MC run finds the native conformation within 5×10^5 MC steps.

The correlation of MFPTs between MS1 and MS2 is shown in Figure 4, which clearly demonstrates that MS2 shows faster folding than MS1. On the average, MS2 is at least 3−6 times faster than MS1 for 100 model protein sequences at $T = 0.45$, although MFPT is strongly affected by the sequence itself.[2,3,11,12] Because MS2 includes not only MS1 itself but also more extended pivot and crankshaft moves which are not possible in MS1, it is not an unexpected result that MS2 shows better foldicity and faster MFPT than MS1. In the process of MC sampling, however, the probability of successful moves in MS2 may be lower than in MS1 because it is less likely that the move of a large fragment does not topologically self-overlap and is energetically acceptable to Metropolis criteria. With all of these considerations, it is a very interesting result that MS2 always resulted in faster folding than MS1. Figure 4 shows that some sequences are slow folding sequences in both MS1 and MS2, even though all the model sequences have been obtained by optimization.

To see how the two parameters in MS2, N_{max_pivot} and N_{max_crank}, affect the MFPT, we performed 100 independent MC runs for 10 model sequences by varying the two parameters, and the result is shown in Figure 5. This figure demonstrates that MFPTs are linearly dependent on these two parameters; that is, the MC lattice folding simulation gets faster when the maximally allowed move size gets bigger. This figure also indicates that MFPT is affected more by the size of the pivot move than that of the crankshaft move.

B. Folding Dynamics of MS1 and MS2. The number of native contacts, Q, is usually used as a reaction coordinate in many lattice folding simulations,[10,11,14] although there are also other useful definitions for reaction coordinates.[13,34,35,40] In this study, we introduce D as a new parameter for reaction coordinates:

$$D = Q - (C - Q)$$

where C is the total number of contacts. In the above expression, the second term $(C - Q)$[36] represents the number of non-native contacts that should be broken first to fold to the native state. Thus, in this study with the $3 \times 3 \times 3$ cubic lattice model, the parameter D can vary from −28 to 28, where the native state corresponds to the state at $D = 28$. The value of D at some state tells us some information about distance measure[13] from the native state, because the value of $Q_{max} - D$, which correspond to the number of native contacts to be formed plus the number of non-native contacts to be broken, may be considered as the distance from the native state.

A typical trajectory of the lattice MC simulation is shown in Figure 6. Here, several points may be mentioned. First, at the initial stage, the total number of contacts, C, rapidly increases, but the value D still remains around zero, meaning that the rapid compaction at the initial phase is not due to the increase of native contacts but due to the random compaction. This result is consistent with the other works previously reported.[2,3,10,13,34,37] Second, MS2 finds the native state around 10^4 MC steps, while MS1 does it around 10^5 MC steps. Finally, after the lattice chain arrives at the native states, or equilibrium state, MS2 behaves more dynamically and moves farther away from the native state than MS1. At $T = 0.45$, the population of the native states in equilibrium is about 55% in MS1 and about 45% in MS2, respectively. This result could be explained by the fact that MS2 allows the move of larger segments by a one-step MC move from the native state, thus requiring a relatively large number of trials to fold back again to the native state.

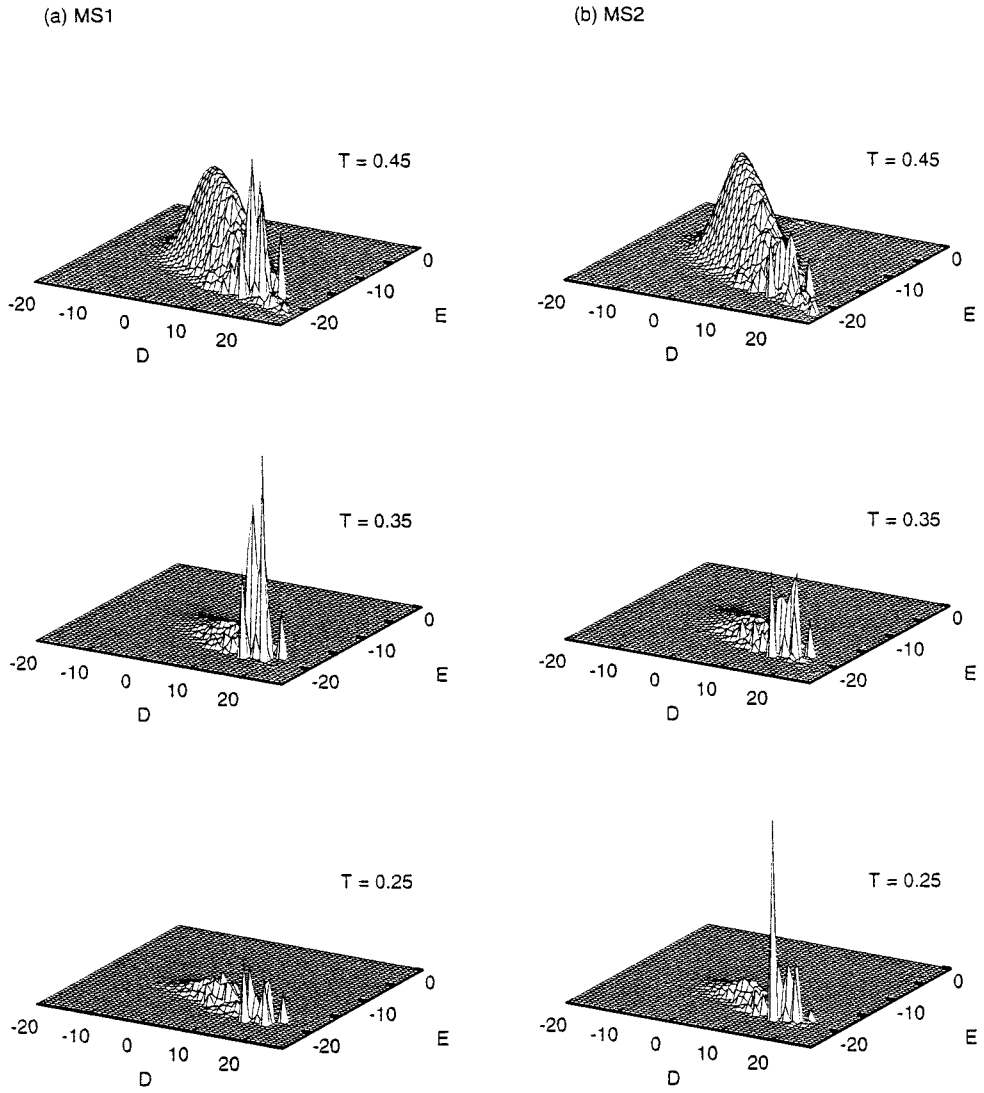

Figure 10. Population densities of MS1 (a) and MS2 (b) at $T = 0.45$, $T = 0.35$, and $T = 0.25$. The population densities are obtained from more than 2000 independent MC folding runs for sequence 32 in Table 1. All the runs are started from random initial conformations and stopped when the native states are found.

The use of D, instead of Q, gives us some advantages in the analysis of lattice folding dynamics. For example, Figures 7 and 8 show the density of the states and the free energy profiles, respectively, of the model sequence 12. When using D as a reaction coordinate, it is clearly seen that there exist some stable intermediates where $D = 16$, while it is not so clear in the plot with Q. In addition, it is readily interpreted that the region where $D < 0$ corresponds to the states that have more non-native contacts than native contacts, but these states cannot be represented by the coordinate Q. Although the use of D as a reaction coordinate clearly loses the exact value of Q, it seems that the parameter D provides more detailed figures than Q, in terms of density of states and folding profiles. It should be noted that the use of both Q and D is merely the crude expression on the true reaction coordinates which should be multidimensional coordinates, as was mentioned in many works.[4,38]

Figures 7 and 8 also show that at equilibrium the conformation spaces sampled by two different move sets, MS1 and MS2, are not very different from each other. As was discussed in ref 39, the overall thermodynamics may not be strongly affected by the use of different move sets. However, since MS2 shows more dynamic moves in equilibrium states as shown in Figure 6, the denatured states, which may correspond to the states where $-5 < D < 5$ or $7 < Q < 13$ in Figures 7 and 8, are sampled more by MS2 than by MS1. Thus, in MS2, the density of states and the free energy for the denatured states are shown to be slightly high and low, respectively.

C. Folding Pathways of MS1 and MS2. Figures 7 and 8, the density of states and the free energy profiles of MS1 and MS2 are obtained by the number of MC runs with 2×10^6 MC steps starting from the native conformation as an initial conformation. Therefore, if we consider that the MFPTs of this sequence at $T = 0.45$ are 9.4×10^4 steps for MS1 and 1.0×10^4 steps for MS2, respectively, the properties derived from these runs may be regarded as the equilibrium properties at corresponding temperature. It appears that the stability and the accessibility of the denatured states are the major differences between MS1 and MS2 in these two figures.

To investigate whether there are some differences in folding kinetics between MS1 and MS2, we carried out long MC runs at two different temperatures and then the folding profiles at any desired Ts are obtained by using the statistical mechanical relation.[36] The results are shown in Figure 9. As shown in Figure 9a, folding profiles obtained from the runs at T close to folding transition T_f are almost the same for MS1 and MS2. However, as shown in Figure 9b, the folding profiles of MS2 obtained from the runs at lower temperature ($T = 0.40$) cannot

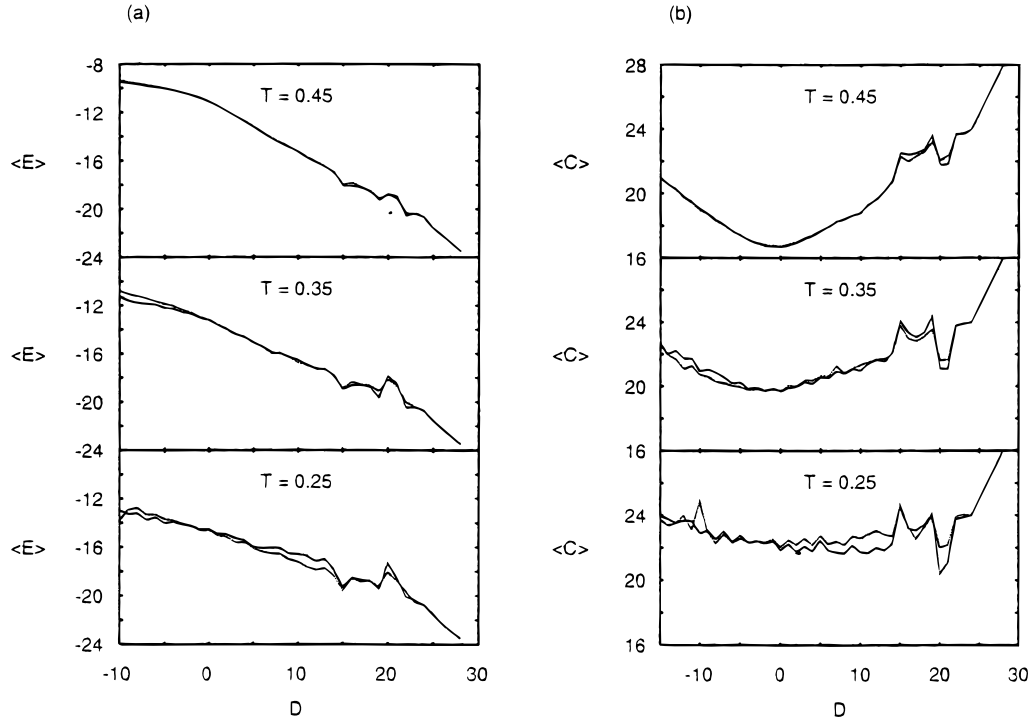

Figure 11. Profiles of energy (a) and the number of contacts (b) plotted against D during folding which are calculated using population data in Figure 10. Solid lines correspond to MS1, and dashed lines correspond to MS2. The energy, $\langle E \rangle$, and the total number of contacts, $\langle C \rangle$, are obtained by average; thus the profiles may correspond to the most probable paths until the model protein finds the native state starting from random initial conformations.

be well matched to those of MS1. Most significant differences between MS1 and MS2 are the free energies for the denatured states. As was noted earlier by other workers,[13] this result shows that the kinetics in the lattice folding simulation are dependent on the move sets and the differences seem to be more pronounced when the simulation temperature is lower than the folding transition temperature. However, since Figure 9 represents thermodynamics rather than kinetics, it should be mentioned that the increased differences at low temperature may be due to the differences in the equilibration times between MS1 and MS2.

More analysis to compare the differences in folding kinetics between the two move sets is possible. One such example is shown in Figure 10, which represents the density of states obtained from more than 2000 MC runs at three different temperatures, $T = 0.45$, $T = 0.35$, and $T = 0.25$, respectively. In this figure, randomly generated, but the same set of, initial conformations for MS1 and MS2 are used for equivalent comparison and each MC run is stopped when the native state is found. At $T = 0.45$, which corresponds to the relatively high temperature for fast folding, a major population is located at the denatured states where $D < \sim 10$, meaning that at $T = 0.45$ the model protein spends a long time in the denatured states until it finds the native state. On the contrary, at lower temperatures, $T = 0.25$ and/or $T = 0.35$, a large population is shown to be shifted to the region toward the native states where $D > 15$. These significantly increased states at lower temperature may be interpreted as the trapped states which have been well explained by the free energy landscape model of protein folding.[4,34,35,40]

The microscopic difference in folding paths between MS1 and MS2 is shown in Figure 10. Comparing to MS2, MS1 shows relatively large population in the states where $15 < D < 23$ at $T = 0.45$, meaning MS1 is more likely to be trapped in stable non-native states located in this region than MS2. But this does not mean that the rate-limiting step is different in MS1 from MS2 at $T = 0.45$ because the major populations in both MS1 and MS2 are still located in the region $D < \sim 10$. At lower temperature, $T = 0.25$, the trapped states where $15 < D < 23$ become the major populations; thus the escape from these stable states seems to strongly affect the folding dynamics in this case. This population shift is more significant in MS2 than in MS1. The size of conformation spaces, in terms of D and E, sampled by the two different move sets is almost the same as shown in Figure 10, meaning that there are no hidden states that are exclusively accessible to one move set. But, as Figure 10 shows, there exists some states that have different accessibility by the move set.

Figure 11 shows the profiles of energy and the total number of contacts obtained from the density of states in Figure 10 plotted against the reaction coordinate D. Several features related with the move sets may be drawn from Figures 10 and 11: (1) The population densities during folding seem to be largely affected by the move sets, as shown in Figure 10. However, the differences between MS1 and MS2 in average energy and average number of contacts shown in Figure 11 are not greater than expected from the density of states in Figure 10. (2) At lower temperature, the differences of the move sets get larger. (3) At $D = \sim 20$, which may correspond to the transition state, MS2 shows more efficient profiles, in terms of energy and total number of contacts, than MS1 because MS2 shows a lower energy profile and keeps a higher number of contacts around $D = \sim 20$ than MS1.

Although it is still not clear whether the move set used in lattice MC simulations really mimics the real protein folding kinetics,[39] our results support the previously reported results[11,13,39] that the move sets clearly affect the microscopic features in lattice MC simulations, especially at low temperature. However, overall thermodynamics at equilibrium condition are likely to be insensitive to the move set.

Conclusions

Using the cubic lattice model, we have designed a more extended move set and carried out comparative studies on the effects of move sets in lattice protein folding. From our study, the following conclusions may be drawn.

(a) In terms of foldicity and MFPT, the MC simulation of lattice folding is highly dependent on the move set used. In the folding simulations for 100 model protein sequences, a more flexible move set (MS2) always results in faster folding and higher foldicities than the commonly used move set (MS1) that has only local moves. This may be because MS2 allows more flexible choices for moves during simulation than MS1. Thus, MS2 may provide a greater advantage in the lattice model study for protein folding.

(b) At equilibrium conditions, although MS2 shows slightly increased population density of denatured states, the overall distribution of the population density of states is almost the same in MS1 and MS2, meaning that the overall thermodynamics may not be strongly affected by the use of different move sets.

Acknowledgment. This work was supported by the STEPI (1997). We thank Drs. H. C. Shin and S. Kang for helpful discussions on this work.

References and Notes

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223.
(2) Karplus, M.; Sali, A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 58−73.
(3) Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561.
(4) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167.
(5) Lau, K. F.; Dill, K. A. *Macromolecules* **1989**, *22*, 3986.
(6) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1990**, *92*, 3118.
(7) Shakhnovich, E. I.; Gutin, A. M. *J. Chem. Phys.* **1990**, *93*, 5967.
(8) Skolnick, J.; Kolinski, A. *Science* **1990**, *250*, 1121.
(9) Hinds, D. A.; Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2536.
(10) Sali, A.; Shaknovich, E. I.; Karplus, M. *Nature* **1994**, *369*, 248-252.
(11) Sali, A.; Shaknovich, E. I.; Karplus, M. *J. Mol. Biol.* **1994**, *235*, 1614−1636.
(12) Miller, R.; Danko, C. A.; Fasolka, M. J.; Balazas, A. C; Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1992**, *96*, 768.
(13) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1994**, *100*, 9238.
(14) Oh, W. S.; Shin, J. M. *Bull. Kor. Chem. Soc.* **1996**, *17*, 808.
(15) Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1282.
(16) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Folding Des.* **1996**, *1*, 221.
(17) Dill, K. A.; Fiebig, K. M.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1942.
(18) Shakhnovich, E. I. *Phys. Rev. Lett.* **1994**, *72*, 3907.
(19) Hao, M. H.; Scheraga, H. A. *J. Phys. Chem.* **1994**, *98*, 4940.
(20) Honig, B.; Cohen, F. E. *Folding Des.* **1996**, *1*, R17.
(21) Shakhnovich, E. I. *Folding Des.* **1996**, *1*, R50.
(22) Metropolis, N.; Rosembluth, A.; Rosembluth, M.; Teller, A. *J. Chem. Phys.* **1953**, *21*, 1087.
(23) Hilhorst, H.; Deusch, J. *J. Chem. Phys*. **1975**, *63*, 5153.
(24) Wetlaufer, D. B. *Proc. Natl. Acad. Soc. U.S.A.* **1973**, *70*, 697.
(25) Karplus, M.; Weaver, D. L. *Nature* **1976**, *260*, 404.
(26) Moult, J.; Unger, R. *Biochemistry* **1991**, *30*, 3816.
(27) Unger, R.; Moult, J. *J. Mol. Biol.* **1993**, *231*, 75.
(28) Camacho, C. I.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6369.
(29) Covell, D. G. *J. Mol. Biol.* **1994**, *235*, 1032.
(30) Nunes, N. L.; Chen, K.; Hutchinson, J. S. *J. Phys. Chem.* **1996**, *100*, 10443.
(31) Goldstein, H. *Classical Mechanics*, 2nd ed.; Addison-Wesley Publishing Company: Reading, MA 1980; Chapter 4.
(32) Myazawa, S.; Jernigan, R. *Macromolecules* **1985**, *18*, 534.
(33) Shakhnovich, E. I.; Gutin, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7195.
(34) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8721.
(35) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626.
(36) Mirny, L. A.; Abkevich, V. I.; Shakhnovich, E. I. *Fold Des.* **1996**, *1*, 103.
(37) Shakhnovich, E. I.; Farztdinov, G.; Gutin, A. M.; Karplus, M. *Phys. Rev. Lett.* **1991**, *67*, 1665.
(38) Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29.
(39) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Chem. Phys.* **1994**, *101*, 6052.
(40) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902.