

# Comparison of three Monte Carlo conformational search strategies for a proteinlike homopolymer model: Folding thermodynamics and identification of low-energy structures

Dominik Gront

*Department of Chemistry, University of Warsaw, Ul. Pasteura 1, 02-093 Warsaw, Poland*

Andrzej Kolinski<sup>a)</sup>

*Department of Chemistry, University of Warsaw, Ul. Pasteura 1, 02-093 Warsaw, Poland and Donald Danforth Plant Science Center, Bioinformatics and Computational Genomics, 893 North Warson Road, St. Louis, Missouri 63141*

Jeffrey Skolnick

*Donald Danforth Plant Science Center, Bioinformatics and Computational Genomics, 893 North Warson Road, St. Louis, Missouri 63141*

(Received 13 April 2000; accepted 27 June 2000)

Entropy sampling Monte Carlo, the replica method, and the classical Metropolis scheme were applied in numerical studies of the collapse transition in a simple face-centered cubic lattice polymer. The force field of the model consists of pairwise, contact-type, long-range interactions and a short-range potential based on the  $\beta$ -sheet definition assumed in the model. The ability to find the lowest energy conformation by various Monte Carlo methods and the computational cost associated with each was examined. It is shown that all of the methods generally provide the same picture of the collapse transition. However, the most complete thermodynamic description of the transition derives from the results of entropy sampling Monte Carlo simulations, but this is the most time-consuming method. The replica method is shown to be the most effective and efficient in searching for the lowest energy conformation. The possible consequences of these findings for the development of simulation strategies for the folding of model proteins are discussed briefly. © 2000 American Institute of Physics. [S0021-9606(00)50836-8]

## I. INTRODUCTION

The last decade has witnessed impressive progress in the development of methodologies for the computer simulation of molecular systems. In particular, a number of new Monte Carlo schemes have been proposed.<sup>1-6</sup> Rather than being just technical improvements, many are new qualitative approaches to the problem of computer simulations. Such computer simulations can be extremely helpful in understanding the complex behavior of biomolecules.<sup>7</sup> For example, the understanding of the molecular mechanism of protein folding is one of the most challenging and urgent tasks of theoretical molecular biology.<sup>8,9</sup> Due to the complexity of such systems,<sup>10</sup> detailed all-atom simulations can cover only a small time interval<sup>11</sup> (which is orders of magnitude shorter than the characteristic folding time for proteins). Thus, it is necessary to employ reduced models.<sup>12-17</sup> The smaller number of explicitly treated degrees of freedom in such models allows the investigation of some aspects of the entire folding process. Of course, as many "essential details" as possible should be included in the reduced models so that the insights that are gained are also applicable to real systems.

In this work we examine a relatively simple polypeptide chain model. The conformational space of the model is re-

stricted to an ensemble of homopolymeric chains located on a face centered cubic (fcc) lattice. The number of allowed rotational isomeric states per chain unit of this model is comparable to the number of conformations per residue in polypeptides. A simple, short-range potential mimics proteinlike local conformational preferences, and the pairwise long-range potential simulates an average hydrophobic attraction between chain units. Thus, it may be expected that some of the most general features of protein chains will be qualitatively reproduced. Due to the relatively large number of conformations and the effects of chain connectivity and packing, the energy landscape of this model may be sufficiently complex to mimic some aspects of the rugged energy surface found in real proteins. When compared to the very popular simple cubic lattice protein models,<sup>12</sup> the present model has several advantages: A larger number of conformations per chain unit, a local geometry that is closer to the geometry of real proteins, a local conformational stiffness that is characteristic of polypeptides, and a coordination number (12) that is qualitatively similar to the average number of side chain contacts per residue in the core of globular proteins. Examinations of various protein sequences having amino-acid-dependent potentials will be addressed in future work. The goal of the present study is to evaluate the applicability of various Monte Carlo (MC) schemes to the problem of protein folding. In particular, we compare three qualitatively different algorithms with respect to both their

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: kolinski@chem.uw.edu.pl

capability of finding the lowest energy state (the “native” state of the protein) as well as the computational cost of obtaining a complete thermodynamic description of the system over the relevant temperature range. The first sampling method we consider is the traditional Metropolis scheme (MS),<sup>18</sup> which is usually used in the context of simulated annealing procedures. The second method examined is the replica method<sup>3</sup> (RM) (or the replica exchange method), in which several independent copies of the system sample conformational space at various temperatures. The third sampling methodology is a version of multicanonical ensemble sampling,<sup>1</sup> or the entropy sampling Monte Carlo (ESMC), scheme.<sup>2,19,20</sup>

The outline of this paper is as follows: For the reader’s convenience, we first provide a brief summary of each of these sampling schemes. Then, we describe the polymer model used in the simulations. The simulation results for these three sampling procedures are used to compare the relative efficiency of these methods and their potential applicability to the protein folding problem. We conclude with a discussion of the implications of the present study to the more general problem of protein folding.

## II. MONTE CARLO SAMPLING METHODS

### A. Metropolis scheme

In the formalism of statistical mechanics, assuming a Boltzmann distribution of states, any physical quantity of a system can be written as follows:

$$\langle A \rangle = Q^{-1} \int A(\mathbf{x}) \exp(-H(\mathbf{x})/k_B T) d\mathbf{x}, \quad (1)$$

where  $A(\mathbf{x})$  is a measurable quantity,  $\mathbf{x}$  represents the coordinates in the conformational space, and  $Q$  denotes the configurational partition function

$$Q = \int \exp(-H(\mathbf{x})/k_B T) d\mathbf{x}, \quad (2)$$

where  $H(\mathbf{x})$  is the Hamiltonian of the system. This formula can be used to estimate various quantities,  $A(\mathbf{x})$ , of a model system by the approach proposed by Metropolis *et al.*<sup>18</sup> In the Metropolis method, a Markov process is constructed, whose unique limiting distribution is the Boltzmann distribution.<sup>21</sup> In such a Markov chain, a new state is generated by a random modification of the preceding one. The change is accepted with a probability  $p(\mathbf{x}_i, \mathbf{x}_{i+1})$ :

$$p(\mathbf{x}_i, \mathbf{x}_{i+1}) = \min\{1, \pi(\mathbf{x}_{i+1})/\pi(\mathbf{x}_i)\}, \quad (3)$$

where  $\pi(\mathbf{x}_i)$  is the Boltzmann probability of state  $\mathbf{x}_i$

$$\pi(\mathbf{x}_i) = \exp(-H(\mathbf{x}_i)/k_B T). \quad (4)$$

Consequently, the average value of  $A$  can be expressed as the simple arithmetic mean

$$\langle A \rangle \cong (1/M) \sum_{i=1}^M A(\mathbf{x}_i), \quad (5)$$

where:  $\mathbf{x}_i$  denotes the coordinates of the  $i$ th state of the Markov chain and  $A(\mathbf{x}_i)$  is the value of the quantity  $A$  observed for this state.

With decreasing temperature, the system can become trapped in a minimum in the energy landscape. Frequently, for a rugged energy surface, this is not the global minimum. At low temperatures, the convergence of the Markov chain to its limiting distribution is very slow. The system is at risk of spending a lot of time in a basin corresponding to the local minimum, or sampling a relatively small region of configuration space, which leaves it wandering between two or more local minima on the energy surface.

### B. Replica method

The numerical study of multiple copies of a model system, simulated in a parallel fashion, has been described by Swedensen and Wang.<sup>3</sup> Generally, it is a composite Markov chain. There are  $N$  separate replicas of the model system and a set of  $N$  different predefined temperatures. Each replica is sampled with the Metropolis scheme as described above. Let us assume that at a given moment of simulation, the  $i$ th replica is associated with the temperature  $T_m$ , and described by the Hamiltonian  $H(\mathbf{X}_i)$ , where  $\mathbf{X}_i$  is the conformation of the  $i$ -th replica. The Hamiltonian has the same form for all copies. Then, the composite Markov chain is constructed as follows: one of the  $N-1$  pairs of replicas ( $i$  and  $j$ —the  $i$ th associated with temperature  $T_m$  and  $j$ th associated with  $T_n$ ), is randomly selected and replicas are swapped with probability  $p_s$  given by

$$p_s = \min(1, \exp(-\Delta))$$

$$\text{with } \Delta = (1/k_B T_n - 1/k_B T_m)(H(\mathbf{X}_i) - H(\mathbf{X}_j)). \quad (6)$$

The  $i$ th replica runs at temperature  $T_m$  and the  $j$ th at  $T_n$ , respectively. Since the exchange probability decreases exponentially with the temperature, only neighboring replicas need to be exchanged. For large temperature differences, transitions between distant replicas can be safely neglected. Swapping two independent replicas moves them into a new region on the energy surface. Thus, the replicas move not only across the conformational space, but also sample various temperatures. At high temperatures, the system easily overcomes energy barriers and, it is believed, uniformly samples conformational space.<sup>22</sup> The replica exchange step should be attempted with a relatively low frequency, allowing for the equilibration of the replicas’ conformations at all temperatures. Copies at low temperature will most likely find local minima of the energy. Finally, the quantity  $A$  for each temperature  $T_m$  can be estimated according to Eq. (1).

### C. Entropy sampling Monte Carlo

With this method, described by Lee<sup>2</sup> and later employed in computer studies of simple proteinlike models,<sup>5,19,20,23–27</sup> the thermodynamic properties of the model system at all temperatures of interest can be obtained from a single simulation. During the Monte Carlo process an entropy-controlled distribution of the system’s conformations is constructed, which enables a straightforward estimation of the system’s entropy as a function of its conformational energy. A trial MC move is accepted or rejected according to probability  $p_s$

$$p_s = \min\{1, \exp[-J(E_{i+1}) + J(E_i)]\}, \quad (7)$$

where  $J(E_i)$  stands for an estimate of entropy for a given energy level  $E_i = H(x_i)$  and  $E_i$  denotes the  $i$ th energy level (in practice,  $E_i$  is a finite energy interval, a bin of a histogram). An estimate of  $J(E)$  is necessary to carry on the simulations. This estimate may be obtained from the following iterative process:

- (i) the initial values of the entropy histogram  $J(E_i)$  are set to 0 for each  $i$ ;
- (ii) A histogram,  $K(E_i)$ , which stores the numbers of conformations at particular energy levels (energy bins)  $E_i$ , is obtained from a sub-run of the ESMC process;
- (iii) A new estimate for  $J(E_i)$  is calculated, according to the following formula:

$$J_{\text{new}}(E_i) = J_{\text{old}}(E_i) + \ln(\max\{1, K(E_i)\}). \quad (8)$$

Steps (ii)–(iii) define a single iteration of the entropy sampling Monte Carlo procedure. During the simulation, the system is “pushed” by the histogram  $J$ , updated in subsequent iterations, into new regions of conformational space until it finally reaches a low-energy state. The process should be repeated until the histogram,  $K$ , becomes flat, i.e., achieves a constant value that is independent of the energy. A flat histogram  $K$  means that the system achieved an artificial distribution of conformations (not the equilibrium Boltzmann distribution), controlled by the transition probability defined in Eq. (7). When converged, ESMC samples all energy levels of the model system with the same average frequency. At that point, the histogram of  $J$  can be treated as an estimate of the entropy of the system:

$$S(E_i) + \text{const.} = J(E_i). \quad (9)$$

The free energy as a function of energy and temperature is available from such a simulation from the formula

$$F(T, E_i) = E_i - TS(E_i). \quad (10)$$

Any physical quantity  $A$  may be computed from the histogram acquired during the simulation by

$$\langle A(T) \rangle = \frac{\{\sum_{i=1}^L \alpha(E_i) \cdot \exp(-F(T, E_i)/k_B T)\}}{\{\sum_{i=1}^L \exp(-F(T, E_i)/k_B T)\}}, \quad (11)$$

where  $L$  is the number of bins in the histogram  $J$ , and  $\alpha(E_i)$  is the average value of property  $A$  for states of energy  $E_i$ .

In the sense that both the average conformational energy and the entropy are obtained from the same simulation, the ESMC method gives the full thermodynamic description of the system. Unfortunately, the method requires a large amount of computer time before it converges.

### III. DESCRIPTION OF THE POLYPEPTIDE MODEL

All of the methods outlined above have been tested on a polymer lattice model restricted to the fcc lattice. The polymer chain consists of  $N$  united atoms (or residues) connected with  $N-1$  vectors. Vector  $\mathbf{v}_i$  connects residues  $i$  and ( $i$

+1) and belongs to the set of twelve lattice vectors of the type  $[\pm 1, \pm 1, 0]$ . The allowed valence angles are 60, 90, 120, and 180 degrees.

To mimic the formation of secondary structure and the conformational stiffness of polypeptides, the following definition of an expanded,  $\beta$ -type chain conformation was implemented. Three subsequent chain vectors are assumed to be in an expanded state when the following criteria are simultaneously satisfied:

- (i) The angles between vectors  $\mathbf{v}_{i-1}$  and  $\mathbf{v}_i$  and between vectors  $\mathbf{v}_i$  and  $\mathbf{v}_{i+1}$  must be greater than 90 degrees;
- (ii) The dot product  $\mathbf{v}_{i-1} \cdot \mathbf{v}_{i+1}$  must be larger than 0.

The short range potential  $U_{i-1,i,i+1}$  depends on three consecutive vectors in the chain  $\mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}$ . For the  $\beta$ -type residues defined above,  $U_{i-1,i,i+1} = -\varepsilon_B$ ; otherwise,  $U_{i-1,i,i+1} = 0$ .

Each residue may have up to twelve neighbors. The long-range potential for two nonbonded chain units is defined as follows:

$$V_{i,j} = \begin{cases} +\infty, & \text{for } r_{i,j} = 0, \\ -\varepsilon_A, & \text{for } r_{i,j} = 1 \text{ (in lattice units),} \\ 0, & \text{for } r_{i,j} > 1 \text{ (in lattice units).} \end{cases} \quad (12)$$

For a chain of length  $N$ , the total energy is the sum of the two contributions

$$E = \sum_{i=2}^{N-1} U_{i-1,i,i+1} + \sum_{i=1}^N \sum_{j \neq i}^N V_{i,j}. \quad (13)$$

Two kinds of local chain modifications were used in all three Monte Carlo processes. The first micromodification involved a randomly selected displacement of the chain ends. The second employed a table of two-bond configurations. An old configuration was substituted with another configuration that fit into the remaining portions of the chain. A single step of the sampling scheme consists of  $N/2$  attempts to make these two-bond moves and two attempts at chain-end moves.

### IV. RESULTS AND DISCUSSION

For purposes of illustration, we have selected a single set of parameters (fixed for all simulations) to describe the model chain. The chain length  $N = 64$  for the fcc lattice (coordination number  $z = 12$ ) has a large number of possible conformations, and the problem of finding the global minimum is nontrivial. However, to check the correctness of the algorithm and convergence of all methods, test simulations were also performed for a smaller system comprised of  $N = 32$  units. The values of the force field parameters ( $\varepsilon_A = 1.0$ ,  $\varepsilon_B = 4.0$ ) mimic the situation of a semiflexible polymer, with a persistence length similar to that estimated for polypeptide chains.

#### A. Application details of the replica method

In order to optimize the replica method, RM, the three following points should be addressed:

- (1) How many replicas must be used?
- (2) What set of temperatures should be used?
- (3) How frequently should the replicas be exchanged?

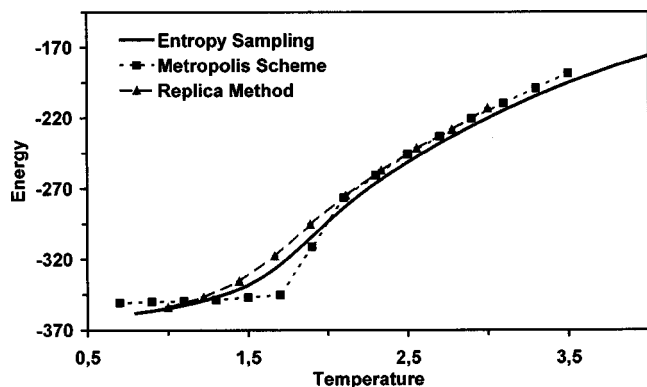


FIG. 1. Energy as a function of temperature from ESMC (thicker solid line), the replica method (dotted line with squares), and the Metropolis scheme (dashed line with triangles).

These parameters are mutually dependent. The wider the temperature range is, the more replicas are necessary. However, a larger number of replicas demands a greater number of replica exchanges. At the same time, the exchanges of replicas should not be too frequent, as the system should be able to relax at its new temperature. Thus, if there are too many replicas, then the cost of finding the global minimum may increase. In order to select an optimal, or reasonable, set of control parameters, the approximate temperature of the phase transition needs to be known. Such an estimation can be obtained from a fast-simulated annealing MS simulation. The selected temperature range should contain the transition midpoint. In the simulations described above, the range of replica temperatures was selected such that the transition temperature was approximately in the center of the range of temperatures sampled.

The temperature difference between replicas need not necessarily be constant. Two types of the replica temperature sets were considered: one having an exponentially changed temperature increment and a linear set with a constant temperature increment. The number of replicas and the temperature range were optimized in a preliminary iterative procedure. A series of short test simulations were performed for various “reasonable” sets of the control parameters. From each run, the lowest observed energy was extracted. When a

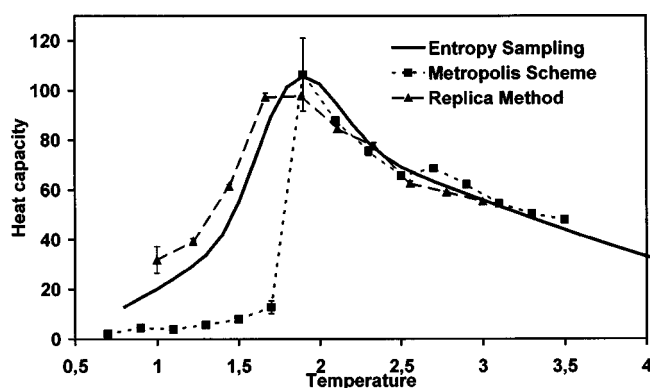


FIG. 2. Heat capacity as a function of temperature from ESMC (solid line), the replica method (dotted line with squares), and the Metropolis scheme (dashed line with triangles).

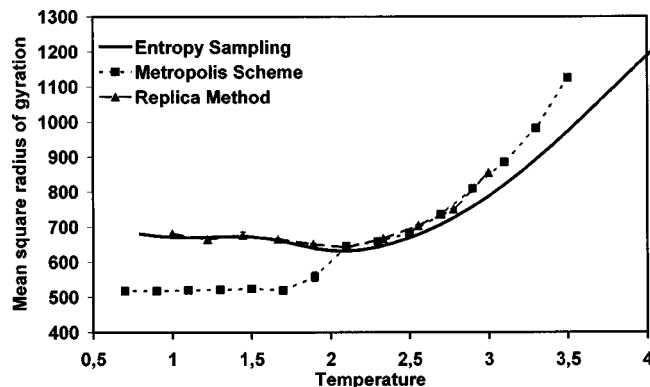


FIG. 3. Mean-square radius of gyration as a function of temperature from ESMC (solid line), the replica method (dotted line with squares), and the Metropolis scheme (dashed line with triangles).

given series of simulations for a given set of control parameters was finished, the average value of the minimal energy was computed. The averages from these various series were compared. The parameters leading to the lowest average value of the lowest energy were selected for the production run. For our model, the following conditions seem to be close to the optimal CPU time needed to reach the lowest energy state:

- (i) Number of replicas: 5
- (ii) Temperature range: 1.25–2.75
- (iii) Frequency of replica exchange: every 1000 steps.

The number of replicas is identical to the number of temperature points at which the system properties can be computed in a straightforward fashion. However, in order to obtain temperature profiles of the various parameters of the system over a wider temperature range, ten replicas were taken and the temperature range was extended to 1.0–3.0.

## B. Collapse transition by various MC procedures

From the ESMC simulations, the entropy of the system as a function of the energy of various states was obtained. The resulting estimate of the partition function enables the calculation of various physical properties, including the average energy,  $E$ , the heat capacity,  $C_V$ , the mean-square-radius of gyration,  $S^2$ , and the percentage of residues in  $\beta$ -type conformations as a function of temperature. The corresponding data have also been obtained using the metropolis scheme and the replica method and the set of results are compared in Figs. 1–4. The error bars for the last two methods were obtained from the numerical data from six independent runs. Since the final estimation of the system’s properties from ESMC was obtained via analytical expressions [see Eq. (11)], the results are continuous and are marked in the plots by thicker solid lines. The accuracy of the results from this method should be the best when convergence of the ESMC method is indeed achieved. While our simulations reached a flat distribution on the energy histogram, suggesting the convergence of the process, some of the lowest energy states seen by the replica method were never visited by the ESMC sampling (see the next section). However, the

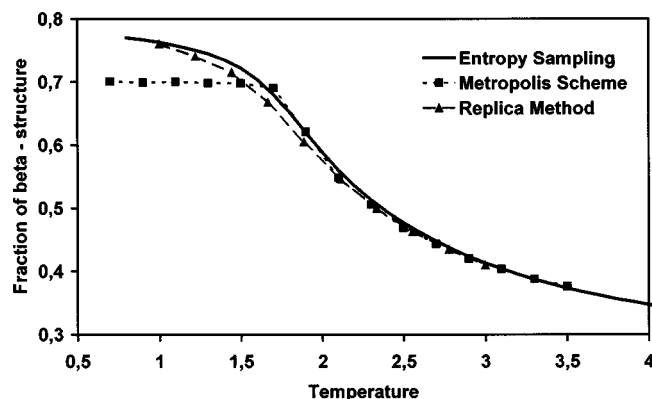


FIG. 4. Fraction of beta-type conformations as a function of temperature from ESMC (solid line), the replica method (dotted line with squares), and the Metropolis scheme (dashed line with triangles).

resulting systematic error of various properties is most likely negligible. This seems to be true even in the very low-temperature range, where the lack of very low-energy states could potentially have important effects on the partition function. Nevertheless, the method is not very efficient at finding the energy minima, at least not in its most straightforward implementation.

All simulations clearly indicate a collapse transition from the expanded random coil state to the dense globular state. This is clearly demonstrated by the plot of the average chain dimensions against temperature (see Fig. 3). The transition is rather smooth. The conformational energy changes gradually (Fig. 1), and the heat capacity peak (Fig. 2) is rather broad. Thus, the cooperativity of the transition is marginal. The collapse transition is accompanied by a large increase of the content of  $\beta$ -type structure. Representative snapshots of the chain conformations are shown in Fig. 5, for  $T=1.25$ , 2.0, and 2.75, respectively. The continuous character of the collapse (or folding) transition can be ascribed to the homopolymeric character of the model polypeptide. At temperatures well below the midpoint of the transition, simple Metropolis simulations (MS) tend to get trapped in the local minima of the energy landscape. This results in a substantial deviation of various estimated properties from the values obtained by both ESMC and the replica method. For ESMC and RM, the problem of local energy barriers is not so acute. ESMC can easily overcome any energy barrier. Likewise, in the replica method, a trapped copy of the system can be exchanged for a new one. As a result, the properties of the system, even at very low temperatures, can be calculated with good accuracy. The results obtained from the replica method have a small systematic error that can be seen in the low temperature range. The average energy is slightly larger than that obtained from ESMC, and the average content of low energy conformations is slightly smaller. This is a result of the “contamination” of average properties at a given temperature by exchanging copies of the system between different temperatures.

As mentioned above, two different temperature sets were compared in the replica sampling protocol: The linear set and the exponential one. The temperature profiles of average energy, heat capacity, and the mean square radius of gyration

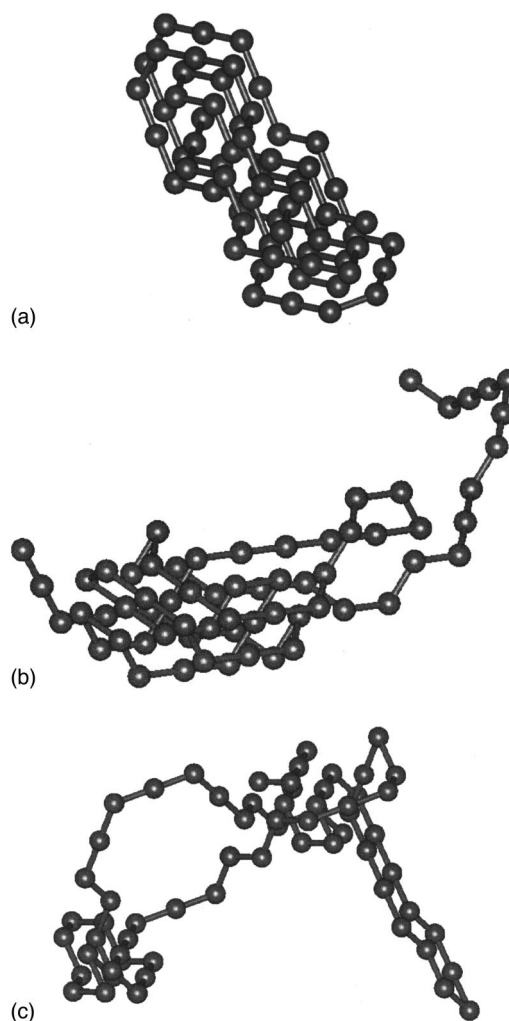


FIG. 5. The snapshots of three example conformations of the model chain obtained by replica method. (a) The most frequently obtained “folded” structure,  $E = -373$ , at  $T = 1.25$ . (b) a representative conformation near the transition temperature,  $T = 1.85$ . (c) an expanded random coil conformation at the temperature well above the transition,  $T = 2.75$ .

were very similar for both sets. Although still in the range of the statistical error, the largest differences were observed for the heat capacity curves, as shown in Fig. 6. Somewhat larger values of the statistical error (for essentially the same simulation time) can be observed for the exponential set of temperatures.

### C. Finding the lowest energy state

The ESMC method is expected to be an excellent method in searching for the lowest energy state. When converged, it finds states whose energy is at least equal to or lower than the best found by the simple Metropolis scheme. Unfortunately, ESMC requires a lot of computer time, and it would be more efficient to perform some MS runs first, in order to generate a conformational pool that can be used to speed up convergence. For short chains, RM was the best protocol in that the lowest energy state for  $N=32$  chains were generated in a much shorter time than when the other two variants of the Monte Carlo method were used. In this case the same energy minimum was found by all sampling

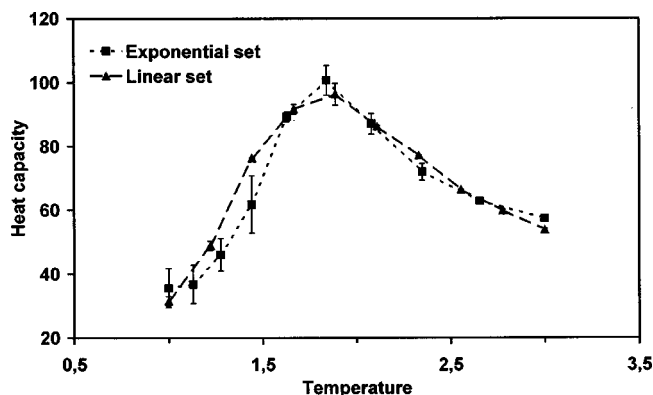


FIG. 6. Heat capacity as a function of temperature, comparison between the linear (dotted line with squares) and exponential temperature set (dashed line with triangles) of replicas in the RM method. In both cases, the simulation time was the same.

procedures. However, for  $N=64$ , which has a more complex search problem, a substantially lower energy state was found by the replica method. The difference in the computational cost of the various schemes is related to the fundamentals of these methods. When the temperature decreases, the classical Metropolis scheme, or its replica implementation, tends to visit a series of states with decreasing average energy, and a vast majority of the high-energy states is neglected. In contrast, ESMC must visit a substantial part of conformational space in order to achieve convergence of the entropy histogram. The longer the chain and/or larger the number of degrees of freedom in the system, the more acute this difference becomes. To partly overcome the problem of slow convergence, a conformational pool, generated in a previous ESMC iteration, can be used to occasionally restart the system trajectory at various energy levels (with a uniform probability).<sup>25</sup> Therefore, the simulation produces more uniform sampling, the entropy barriers are easily surmounted, and convergence is greatly improved.

For the replica method and Metropolis sampling, minimal energy values obtained from ten independent experiments were examined. The replica method found the lowest energy conformation. The results of the search for the lowest energy state by various methods are compared in Table I. The table also provides a comparison of the cost of computations on a 500 MHz Pentium II processor. Two types of globular structures in the low-energy region were found for this model. The first is a seven-stranded  $\beta$ -barrel with an

energy of  $-373$ . A number of very similar structures with the same energy ( $E=-373$ ) were observed in the simulations. The second type of the low-energy structures ( $E=-374$ ) resemble an elongated torus or an 8-stranded  $\beta$ -barrel, with a small opening in the center (see Fig. 7). This structure better optimizes the short-range interactions (characterized by a lack of narrow turns) for the cost of somewhat worse packing. Implementation of different flexibility or secondary structure propensities along the chain should break this kind of degeneracy.

Did any of these methods find the lowest energy conformation for the system? There is no proof that this is the case. Nevertheless, the fact that the states with  $E=-373$  and  $E=-374$  were visited by the simulation process quite often suggests that the simulations do reach the lowest energy states. It is worth noting that for a smaller system ( $N=32$  and the same interaction scheme) all three sampling methods (including ESMC) detected the same lowest energy state.

In summary, the replica method finds much lower energy states (possibly the lowest) than the two other methods for a comparable amount of computer time. The difference of  $10 k_B T$  in the system's energy (see Table I) has a qualitative meaning. The number of states (that were never visited by ESMC or MS) in this range of energy is large. For more complex models of proteins, the discussed differences of the performance of various MC methods might even be more dramatic.

## V. CONCLUSIONS

In this series of simulations, we demonstrated that the replica method is much faster and more accurate than the classical metropolis scheme in finding the energy minima. The most complete estimation of the system's properties can be achieved by the ESMC method. This is because the method provides a straightforward measure of the entropy and energy over the entire relevant range of temperatures. In contrast to the Metropolis scheme or the replica method, entropy Monte Carlo sampling is quasi deterministic—subsequent iterations provide a better estimation of the entropy. As observed in the energy plot (Fig. 1), the average energy obtained from the replica method is slightly higher than the energy derived from the ESMC method. Due to the replica exchange process, small systematic errors occur.

ESMC provides a description of the system's thermodynamics over the entire range of temperatures. However, the

TABLE I. Comparison of the simulation times and ability to find the low-energy states for classical Monte Carlo, RM, and ESMC search schemes.<sup>a</sup>

Method	Temperature set	Number of iterations	Computer time	Average of ten runs of the minimum energy (standard deviation)	Lowest observed energy
MS	linear	$1.1 \cdot 10^8$	1 h 57 min	$-349.3 (\pm 2.067)$	-362
RM	linear	$10^8$	2 h 14 min	$-368.2 (\pm 0.783)$	-373
RM	exponential	$10^8$	2 h 20 min	$-369.7 (\pm 0.789)$	-374
ESMC	n.a.	$\sim 10^9$	$\sim 20$ h	-364	-364

<sup>a</sup>Ten independent simulations were performed for classical Metropolis sampling, MS, and replica sampling, RS, while only one independent simulation was done for the entropy sampling Monte Carlo, ESMC sampling scheme.

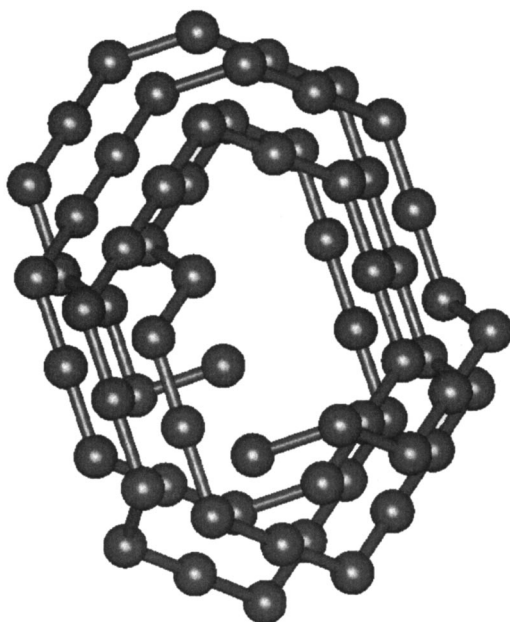


FIG. 7. The lowest energy state,  $E = -374$ , obtained by the replica method at a temperature of 1.25.

convergence of ESMC is too slow when the goal is finding the lowest energy state. The replica method gives good estimations of the system parameters over a wide range of temperatures in a reasonable amount of CPU time. It reaches the basin of low energy states in the shortest CPU time. Thus, RM seems to be the most useful tool for minimization. Interestingly, the two sets of replica temperatures—linear and exponential—compared in this work led to very similar results.

In the forthcoming work, we will apply Monte Carlo methods to find a minimal model that reproduces the most essential features of globular proteins, i.e., a unique structure of the folded state and a cooperative, all-or-none folding transition. A number of sequences will be investigated, and the replica method will be used to find its lowest energy conformations. For those sequences that have a unique ground state, the other parameters of the model will be optimized to reproduce an all-or-none folding transition. The

folding thermodynamics will be investigated in detail by the ESMC method.

## ACKNOWLEDGMENTS

This research was supported in part by Grant No. GM-37408 of the Division of Medical Sciences of the NIH and in part by KBN (Poland) Grant No. 6PO4A-1413.

- <sup>1</sup>B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1991).
- <sup>2</sup>J. Lee, *Phys. Rev. Lett.* **71**, 211 (1993).
- <sup>3</sup>R. H. Swedens and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- <sup>4</sup>H. A. Scheraga, M.-H. Hao, and J. Kostrowicki, in *Methods in Protein Structure Analysis*, edited by M. Z. Atassi and E. Appella (Plenum, New York, 1995).
- <sup>5</sup>H. A. Scheraga and M.-H. Hao, *Adv. Chem. Phys.* **105**, 243 (1999).
- <sup>6</sup>U. H. E. Hansmann and Y. Okamoto, *J. Comput. Phys.* **18**, 920 (1997).
- <sup>7</sup>U. H. E. Hansmann and Y. Okamoto, *Curr. Opin. Struct. Biol.* **9**, 177 (1999).
- <sup>8</sup>A. Sali, E. Shakhnovich, and M. Karplus, *Nature (London)* **369**, 248 (1994).
- <sup>9</sup>O. B. Ptitsyn, *J. Protein Chem.* **6**, 273 (1987).
- <sup>10</sup>R. Zwanzig, A. Szabo, and B. Bagchi, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 20 (1992).
- <sup>11</sup>C. L. I. Brooks, M. Karplus, and B. M. Pettitt, *Adv. Chem. Phys.* **71**, 1 (1988).
- <sup>12</sup>K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, *Protein Sci.* **4**, 561 (1995).
- <sup>13</sup>M. Levitt, *Curr. Opin. Struct. Biol.* **1**, 224 (1991).
- <sup>14</sup>A. Kolinski and J. Skolnick, *Lattice Models of Protein Folding, Dynamics and Thermodynamics* (R. G. Landes, Austin, TX, 1996).
- <sup>15</sup>A. Kolinski, P. Rotkiewicz, and J. Skolnick, in *Monte Carlo Approaches to Biopolymers and Protein Folding*, edited by P. Grassberger, G. T. Barkema, and W. Nadler (World Scientific, Singapore, 1998), pp. 110–130.
- <sup>16</sup>J. Skolnick and A. Kolinski, *Adv. Chem. Phys.* **105**, 203 (1999).
- <sup>17</sup>J. Lee, A. Liwo, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2025 (1999).
- <sup>18</sup>N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **51**, 1087 (1953).
- <sup>19</sup>U. H. E. Hansmann and Y. Okamoto, *J. Comput. Chem.* **14**, 1333 (1993).
- <sup>20</sup>M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 4940 (1994).
- <sup>21</sup>K. Binder, *Monte Carlo Methods in Statistical Physics* (Springer-Verlag, Berlin, 1986).
- <sup>22</sup>K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- <sup>23</sup>M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 9882 (1994).
- <sup>24</sup>M.-H. Hao and H. A. Scheraga, *J. Chem. Phys.* **102**, 1334 (1995).
- <sup>25</sup>A. Kolinski, W. Galazka, and J. Skolnick, *Proteins* **26**, 271 (1996).
- <sup>26</sup>A. Kolinski, W. Galazka, and J. Skolnick, *J. Chem. Phys.* **108**, 2608 (1998).
- <sup>27</sup>D. Mohanty, A. Kolinski, and J. Skolnick, *Biophys. J.* **77**, 54 (1999).