

# A new combination of replica exchange Monte Carlo and histogram analysis for protein folding and thermodynamics

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, 02-903 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 20 December 2000; accepted 3 May 2001)

A novel combination of replica exchange Monte Carlo sampling techniques with a histogram analysis approach is developed and applied to study the thermodynamics of the folding transition in a face-centered cubic lattice chain protein model. Sequences of hydrophobic (*H*) and polar (*P*) topology residues were designed to fold into various  $\beta$ -barrel type proteins. The interaction scheme includes the short-range propensity to form extended conformations, residue-dependent long-range contact potentials, and orientation-dependent hydrogen bonds. Weakly cooperative folding transitions could be observed for properly designed *HP*. (Hydrophobic and polar residue sequences with cooperative long-range interaction methods were proposed and tested.) Based on the study of these systems, the computational cost of such an approach is many times less than the cost of other Monte Carlo procedures. This opens up the possibility for efficient studies of the folding thermodynamics of more detailed protein models. © 2001 American Institute of Physics.  
[DOI: 10.1063/1.1381062]

## I. INTRODUCTION

In a recent paper,<sup>1</sup> we compared the computational efficiency of three distinct Monte Carlo strategies for their ability to search the conformational space of proteinlike semiflexible lattice polymers. These simulations indicated that the Replica Exchange Monte Carlo (REMC) method<sup>2-5</sup> finds the lowest energy state much faster than the simple simulated annealing METROPOLIS<sup>6,7</sup> scheme (MC). REMC is also much faster than the generalized ensemble<sup>8-10</sup> method or its version known as the Entropy Sampling Monte Carlo (ESMC) method.<sup>11-13</sup> However, when successfully converged, ESMC simulations provide a complete thermodynamic description of the model system over the relevant temperature range.<sup>14-17</sup> In this work, we demonstrate that it is possible to extract complete thermodynamics from very fast and efficient REMC simulations. This was achieved by combining histogram methods<sup>18,19</sup> with the REMC method. The ESMC method was used to check the accuracy of the entropy estimation from the REMC results.

In the REMC method,<sup>2-5</sup> several copies of the system are simulated at various temperatures using a standard version of the METROPOLIS scheme. Occasionally, conformational energies of the system's replicas are compared and replicas are swapped according to a probabilistic criterion dependent on energy and temperature differences. Thus, the

model system samples not only conformational space but also various temperatures. A replica trapped at a local minimum of the energy landscape at low temperature has a chance to be moved to a higher temperature, where energy barriers are easier to overpass.

The ESMC method,<sup>11</sup> a version of the multicanonical Monte Carlo algorithm,<sup>8-10</sup> was used here to check our estimation of the system's free energy by a combination of REMC and histogram techniques. The search of conformational space by means of the ESMC technique is controlled by an estimate of the system entropy. An artificial distribution of states is generated, where all energy levels (when converged) are visited with the same frequency. Details of the present implementation of the ESMC method and the REMC method can be found in our previous publication.<sup>1</sup>

The model test system used here is similar to the one previously employed.<sup>1</sup> The lattice representation of the model chain is identical. However, there are qualitative differences between the two models. In this work, we study a heteropolymeric model with *HP* (hydrophobic and polar residue) sequences,<sup>20</sup> mimicking the amino acid patterns in  $\beta$ -type globular proteins. An additional important modification is the introduction of explicit cooperative interactions that qualitatively simulate the averaged effect of main-chain hydrogen bonds in proteins.

Thus, the purpose of this contribution is twofold. First, we describe a modification of existing Monte Carlo protocols that leads to fast identification of the lowest energy state

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

while providing an accurate description of the system's thermodynamics. Second, we assess this improvement in computational sampling, describe the behavior of a minimal model of proteinlike polymers, and discuss the possible consequences of our studies for more detailed modeling of protein structure and thermodynamics.<sup>21</sup>

## II. COMBINATION OF THE HISTOGRAM METHOD WITH THE REMC SAMPLING TECHNIQUE

In a canonical ensemble at temperature  $T$ , the distribution of the system's conformational energy  $E$  obeys a Boltzmann distribution:

$$P(E) = Z(T)^{-1} \cdot w(E) \exp(-E/kT), \quad (1)$$

where  $w(E)$  denotes the density of states and  $Z(T)$  is the partition function at this temperature. Multiplying the distribution by  $\exp(E/kT)$  and calculating the logarithm, the relative entropy (as a function of conformational energy) can be expressed as follows:

$$S(E) = \log(w(E)) = \log(P(E)) + C(T) E/kT, \quad (2)$$

where  $C$  is a constant dependent on the temperature [ $\ln Z(T)$ ] and  $k$  is Boltzmann's constant.

In the histogram method proposed many years ago by Ferrenberg and Swendsen,<sup>18,19</sup> an isothermal MC trajectory is used to estimate the energy distribution  $P(E)$  in the form of a histogram. In a single isothermal simulation, the distribution of states is usually quite narrow. Therefore, a dependable estimation of the entropy can be obtained only over a relatively narrow energy range. However, we show here that overlapping histograms acquired at different temperatures could be combined. Bins of the histograms corresponding to the same energy level provide entropy estimations that differ by a constant value that depends on the temperature-dependent  $Z(T)$ . Since the constants depend only on temperature, they can be subtracted from the two histograms. In practice, the combination of two entropy curves (or rather histograms) needs to be done by a "best fitting" of the overlapping portions of the histograms. This way, several entropy curves calculated from different temperatures may be combined into a single curve that covers a wide range of energies.

In this work, we propose to use histograms,  $P(E)$ , collected during the REMC simulations at various temperatures. Of course, the tails of histograms collected in isothermal MC simulations deviate (sometimes significantly) from the Boltzmann distribution. This is also true for the data generated within the REMC scheme (see Fig. 1). Additionally, due to the exchange of replicas among different energy levels, REMC sampling may potentially introduce some systematic errors in the tails of energy histograms. On the other hand, the exchange of replicas can move the system between various energy minima at low temperatures. As a result, the histograms from REMC at low temperatures are actually more dependable than from isothermal MC simulations. The smallest error should be expected for points (or bins) of energy close to the equilibrium energy for a given temperature. Indeed, the curves fit each other well except for the tails, as demonstrated in Fig. 2. Therefore, the data need to be fil-

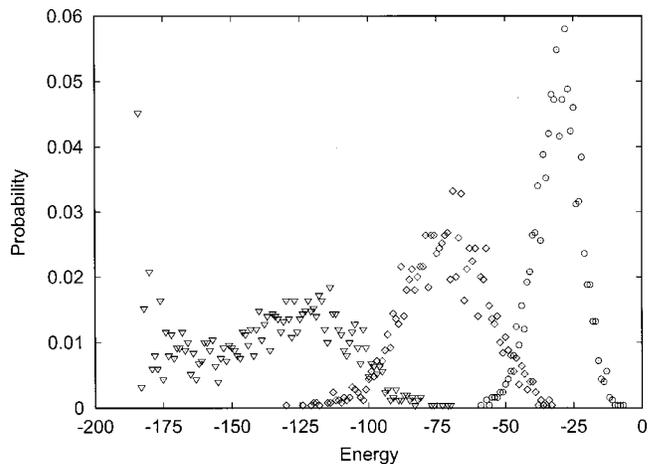


FIG. 1. Three histograms of energy obtained for the S2 sequence with  $\epsilon_c = -2$  (see the text for explanations), for three selected temperatures:  $T = \infty$  (circles),  $T = 3$  (diamonds), and  $T = 1.875$  (triangles). The folding transition temperature is  $T_f = 1.821$ . The replica exchange Monte Carlo method employed a larger number of temperature levels (20 or 30 replicas).

tered. We propose the following, rather conservative procedure.

First, we remove from the energy histogram all points for which the  $P(E)$  value is smaller than  $1/X_{\text{cutoff}}$  of the maximum value seen in the histogram. The value of  $X_{\text{cutoff}}$  was assumed to be in the range of 100. This filter removes uncertain points, which were obtained from a relatively small number of counts. Next, the corresponding entropy histogram is calculated and approximated by a polynomial. Then the points for which the values of  $P(E)$  differ more than  $Y_{\text{cutoff}}$  from their analytical approximations are rejected. The value of  $Y_{\text{cutoff}}$  is of the same magnitude as the error in the entropy approximation by a polynomial (in the present case range of 0.3–0.8). We found that a polynomial of the fifth-order fits the data very well with a correlation coefficient larger than 0.9999. The largest deviations from the Boltzmann distribution were observed at the lowest temperatures,

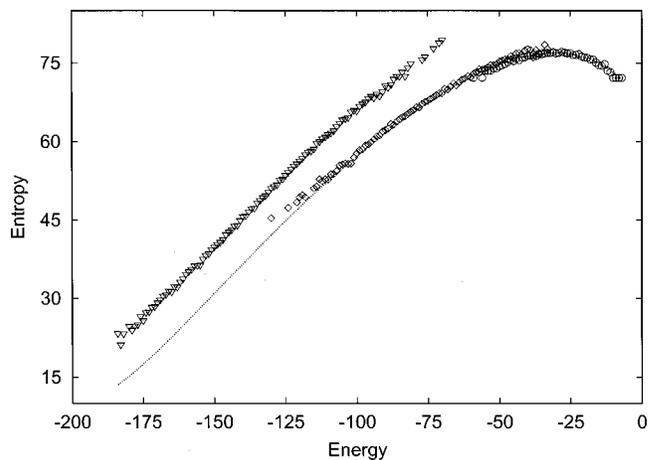


FIG. 2. Relative entropy computed for the histograms given in Fig. 1. The solid line represents the combined entropy curve after the data filtering procedure described in the text. The symbols have the same meaning as in Fig. 1.

where the replicas tend to be periodically trapped in local (or global) minima of the energy.

After the filtering procedure, the entropy curves were sequentially combined, starting with a pair of histograms at relatively high temperatures, where the accuracy of the distribution of energies was expected to be the best. The remaining portions of the entropy curve could be combined one-by-one into a common curve. The shift values were taken from the difference of the values of the corresponding polynomials at the point of largest overlap of the energy histograms. The proposed method of combining data from various temperatures is computationally less demanding than the original histogram method, where the relaxation time for various temperatures needed to be estimated.<sup>18,19</sup>

### III. DESCRIPTION OF THE PROTEIN MODEL

The model polypeptide chain consists of  $N$  residues (beads) connected by  $N-1$  virtual bonds. The chain beads are restricted to the nodes of a face-centered cubic (fcc) lattice. Thus the chain bond vectors belong to the set of 12 fcc lattice vectors of type  $|\pm 1, \pm 1, 0|$ . The model polypeptide contains two types of residues, hydrophobic ( $H$ ) and polar ( $P$ ). The sequence-dependent, long-range interactions are considered only for the nonbonded nearest neighbors on the lattice. There are three values of pairwise interactions:  $\epsilon_{PP}$ ,  $\epsilon_{HH}$  and  $\epsilon_{HP}$ .

Additional strongly directional potentials of long-range interactions mimic the averaged effect of the main-chain hydrogen bonds in proteins. Two residues,  $i$  and  $j$ , are considered to be bonded by a hydrogen bond when the following geometrical criteria are satisfied:

- (1) The interacting fragments of the chains are parallel (or antiparallel), that is, the chain vector  $\mathbf{v}_i$  (connecting the  $i$ th residue with the  $i+1$ th residue) and vector  $\mathbf{v}_j$  are parallel.
- (2) The hydrogen bond vector  $\mathbf{h}_{i,j}$  is orthogonal to  $\mathbf{v}_i$  and to  $\mathbf{v}_j$ .

It is assumed that the hydrogen bonds can only be formed between like residues ( $HH$  or  $PP$ ) and that the energy of hydrogen bonds is constant and equal to  $\epsilon_{\text{Hbond}}$ . The above-mentioned geometrical conditions imply that a given  $\beta$  strand can be fully "hydrogen bonded" to at most two other  $\beta$  strands (although the lattice allows for six contacting parallel strands). This way, the most general feature of the  $\beta$  sheets is qualitatively reproduced.<sup>22</sup> The hydrogen bond network of the model is cooperative in an explicit fashion.<sup>15,21,23</sup> Namely, an additional energy gain  $\epsilon_C$  is assumed for the situations when the  $i$ th residue creates a hydrogen bond with the  $j$ th residue and when simultaneously the  $i+1$ th residue creates a hydrogen bond with the  $j+1$ th residue (or with the  $j-1$ th, residue when the directions of the chain fragments are opposite).

The short-range potential that mimics the conformational propensity toward formation of an extended set of  $\beta$  strands is the same as in our previous work. An energy  $\epsilon_{\text{beta}}$  is associated with all residues for which the three subsequent chain

vectors belong to a set of extended conformations (as defined in Ref. 10). The  $\beta$ -sheet propensity is assumed to be sequence independent.

Two kinds of local modifications were used to mimic the dynamics of the model chain. The first is a random displacement of the two-bond fragments at the randomly selected chain ends. The moves of the inner bonds employ a table of all possible two-bond configurations. The residue index and the new local conformation were selected by a pseudorandom mechanism. The same conformational updating scheme was used in the entropy sampling and replica exchange Monte Carlo simulations.

### IV. DETAILS OF SIMULATIONS

A set of 20 temperatures was used in the REMC simulations. To define an appropriate range for these temperatures, an estimation of the collapse transition temperature is needed, and was obtained from a short simulated annealing run. The lowest temperature for the replicas should not lie much below the transition point because of the freezing of the chain's motion. Most of the replicas were placed in equal temperature intervals around the collapse transition. Three very high temperature replicas were added to the set with  $T=10, 1000, \text{ and } \infty$ . The high temperatures are needed to ensure an unbiased estimation of the histograms for high values of the average energy of the system. Additional high temperature replicas remove the bias for the lower (but still well above the transition) temperatures.

Having the estimation of the system entropy as a function of energy  $S(E)$ , the system's free energy can be calculated,

$$F(E, T) = E - T \cdot S(E). \quad (3)$$

All other properties of the system can be calculated using the partition function defined by the  $F(E, T)$  dependence.

The entropy sampling Monte Carlo (ESMC) technique can be used to test the entropy and free energy calculations obtained with the scheme proposed in this work. This can be done in two somewhat different ways. First, one may use the entropy estimation from the REMC simulations as a starting point for the ESMC. For the true entropy  $S(E)$  curve, the convergence criterion (a flat histogram of the frequency of the system's visits to various  $E$  bins) of the ESMC method should be instantly satisfied. Alternatively, one may construct the entropy estimation from scratch, performing a full set of ESMC iterations (see the previous work for details). The second approach is somewhat more conservative. Both procedures were used to test the REMC-based estimations of the system's entropy. A conformational pool extracted from the REMC "trajectories" was employed to speed-up the convergence of the ESMC procedure. The conformational pool contained a number (the same for each energy bin) of conformations for each energy level accessible for the model system. During the ESMC simulations, the randomly selected conformations from the pool were occasionally (but rarely) used to restart the sampling process. Thus, the problem of entropic barriers (that can significantly slow down the convergence of the ESMC process) can, to a large extent, be overcome.

The REMC and the test ESMC simulations were done for the following parameters of the model interaction scheme,

$$\epsilon_{HH} = -2,$$

$$\epsilon_{PP} = -1,$$

$$\epsilon_{HP} = 0,$$

$$\epsilon_{\text{Hbond}} = -2,$$

$$\epsilon_{\text{beta}} = -4,$$

$$\epsilon_C = 0 \text{ or } -2 \text{ (both cases were tested).}$$

Long trajectories ( $10^8$  attempts to chain modification per replica) for the different model chain sequences were generated by REMC. Attempts to change replicas were executed every 1000 steps. The computational cost was about 6 h CPU on a fast PC.

The system entropy was also calculated from the ESMC simulations. These results were compared with the REMC estimation of the entropy. The test of the REMC convergence (using the entropy curve as an input for ESMC) by ESMC cost 3.5 CPU hours. An independent (entropy generated from scratch) ESMC run took about 70 h.

## V. RESULTS AND DISCUSSION

### A. Topology of the lowest energy states

Three *HP*-type sequences were tested. The design of the first sequence, S1, was targeted on the simple up-and-down topology of the lowest energy state. The second and the third sequences that have a possibility to form a longer loop may lead to formation of the Greek key topology,<sup>22</sup>

$$S1:(HPHPH)_6,$$

$$S2:(HPHPH)_3 \text{ HPHPHPP}(HPHPH)_2,$$

$$S3:(HPHPHPHPH)_3$$

$$HPHPHPHPHPP(HPHPHPHPH)_2.$$

For the first sequence, S1, a single nondegenerate structure (modulo, the topological mirror image) of the lowest energy was found. It was indeed the up-and-down six-member  $\beta$  barrel. Each strand consisted of five residues, with three strands per sheet. The snapshot of the folded structure is shown in Fig. 3. For the second, S2, sequence, three conformations at the lowest energy bin were observed. All of them were six-member  $\beta$  barrels. A small fraction of the folded structures had an up-and-down topology similar to that of S1. The second type of lowest energy structure had the Greek key topology (a common motif in globular proteins). The third structure had some features of the Greek key, however it is incomplete (Fig. 4). Interestingly, the value of the cooperativity parameter ( $\epsilon_C = 0$ , or  $-2$ ) had no influence on the folded structures. For the nonzero cooperativity, the energy of the ground state structures was lower. This sequence (S3) has the same pattern of *HP* residues as the S2 sequence. Longer strands were designed to find out if their length has any influence on the folding thermodynamics.

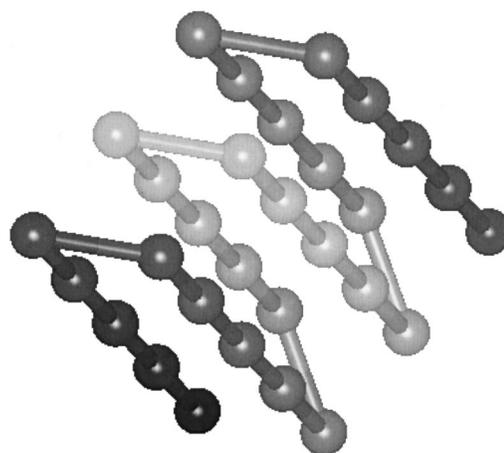
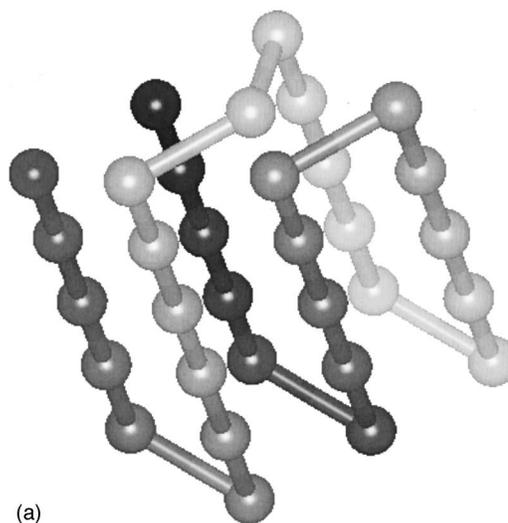
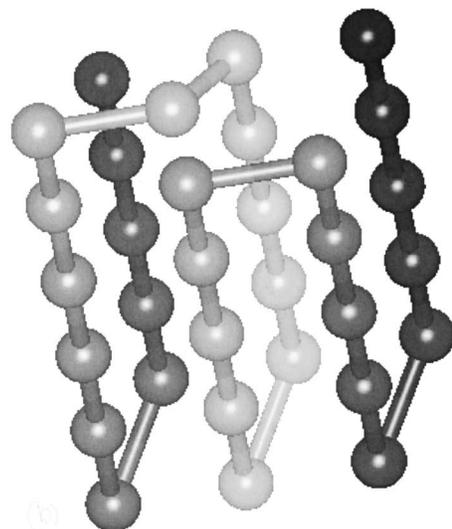


FIG. 3. The lowest energy conformation for the S1 *HP* chain. The mirror image conformation has the same contact map and the same total energy.



(a)



(b)

FIG. 4. The lowest energy conformations of the S2 sequence. (a) Greek key fold. (b) A permutation of the Greek key fold. The third competing structure (seen rarely) has the up-and-down conformation similar to that shown in Fig. 3.

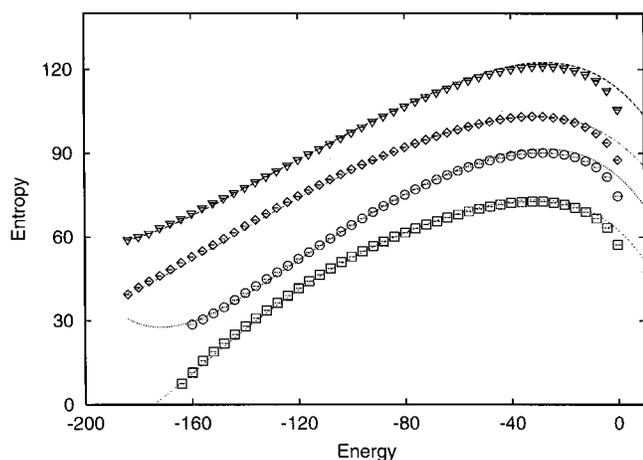


FIG. 5. Entropy as a function of conformational energy for four model systems. Symbols denote the test results from ESMC with: triangles for S1 and  $\epsilon_C = -2$ , diamonds for S2 and  $\epsilon_C = -2$ , circles for S1 with  $\epsilon_C = 0$ , and squares for S2 with  $\epsilon_C = 0$ . In all cases, the error is smaller than the symbol size. The lines are polynomial approximations of the REMC results.

## B. Folding thermodynamics

Our combination of the replica exchange Monte Carlo sampling with the histogram method provides a quite complete thermodynamic description of the model system over a wide range of temperatures. In contrast to other MC techniques, the present application of the REMC enables a simultaneous estimation of the system's conformational energy and entropy from a single simulation. In Fig. 5, the entropy calculated from the REMC is compared with the entropy computed by means of the ESMC method. It is clear that the two methods give essentially the same results, except for the (rather irrelevant) region of the highest values of the energy. The free energy can then be calculated by Eq. (3) for any temperature of interest. The  $F(E, T)$  dependence (we use the data approximated by a polynomial) enables the calculation of any property of the system. In particular, the average conformational energy can be calculated as a function of temperature as follows:

$$\langle E(T) \rangle = \frac{\sum \{E_i \exp(-F(E_i, T)/kT)\}}{\sum \{\exp(-F(E_i, T)/kT)\}}, \quad (4)$$

where index  $i$  enumerates particular bins of the energy histogram.

In Fig. 6 the mean energy calculated from Eq. (4) is compared with the average energy of replicas running at given temperatures (the arithmetic mean of the energy at a given temperature). Good agreement between the two types of data provides a strong additional test of the proposed combination of the REMC and histogram methods.

Free energy curves provide information on the possibility of the existence of phases (or states) in equilibrium (Fig. 7). Only for the S2 and S3 sequences (designed for the Greek key topology) with  $\epsilon_C = -2$  were weakly cooperative folding transitions observed. The height of the free energy barrier is about 1 kT when estimated from REMC data, and 0.5–1 kT when calculated by the ESMC method as shown in Fig. 7. This is slightly above the error of the method. In the remain-

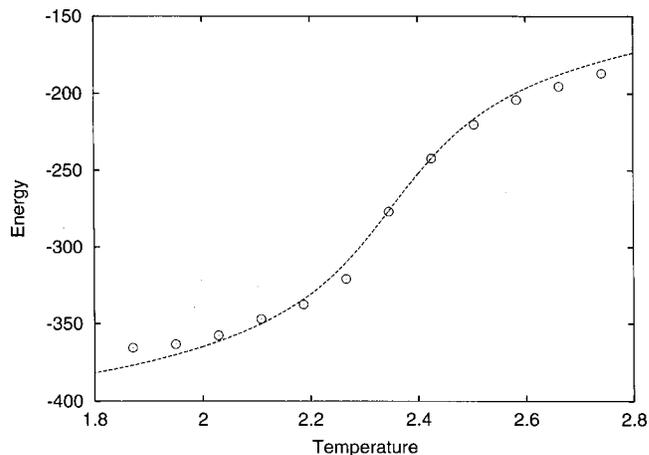


FIG. 6. Comparison of the mean energy as a function of temperature for the S3 sequence with  $\epsilon_C = -2$ . Symbols stand for the straightforward averages for various temperature levels in REMC simulations. The solid line represents the data from the histogram method calculated according to Eq. (4).

ing cases the  $F(E, T)$  curves do not indicate any "phases" at equilibrium. The folding transition has a continuous character for these systems. These results suggest that the cooperativity of the folding transition depends on sequence and

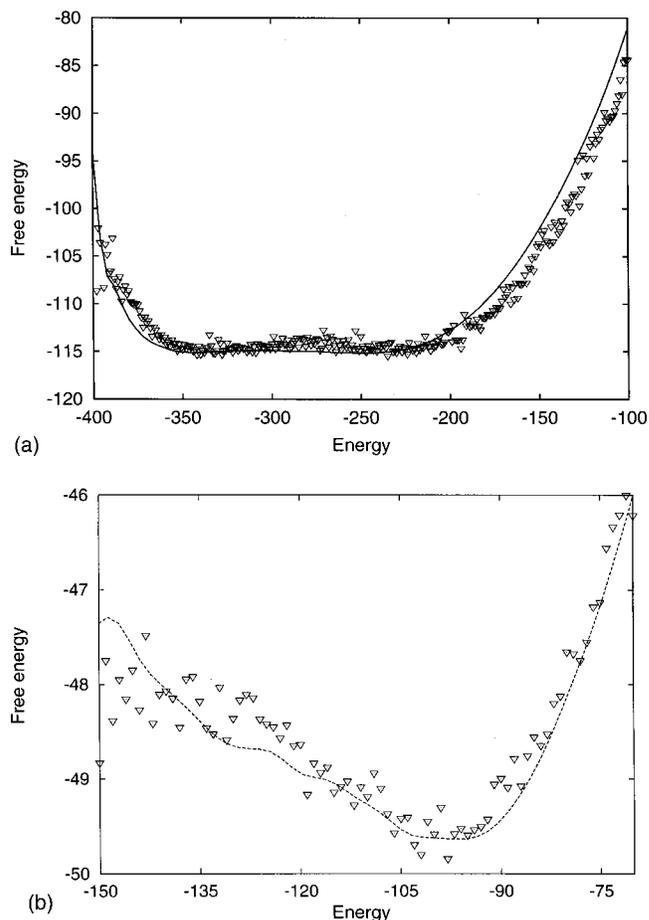


FIG. 7. Free energy as a function of energy for two sequences at the folding transition temperature. The lines are for the ESMC data, the triangles are for the REMC with histogram method. (a) For the weakly cooperative S3 ( $\epsilon_C = -2$ ) system, the transition temperature is  $T_f = 2.334$ . (b) Sequence S1 ( $\epsilon_C = 0$ ) exhibits a continuous transition at  $T_f = 1.671$ .

cooperative long-range interactions (hydrogen bond interactions), but not on the chain length. In all of the cases, the transition temperatures were estimated from the energy fluctuations and calculated from a formula similar to Eq. (4). The differences are in the range of 0.04.

### C. Comparison with a homopolymeric model

It is interesting to compare the behavior of the *HP* model with an otherwise similar homopolymeric model studied in our recent work.<sup>1</sup> In terms of this paper, the homopolymeric sequence S4 is  $(H)_{30}$  with the following force field parameters:

$$\epsilon_{HH} = -1,$$

$$\epsilon_{PP} = 0,$$

$$\epsilon_{HP} = 0,$$

$$\epsilon_{\text{Hbond}} = 0,$$

$$\epsilon_{\text{beta}} = -4,$$

$$\epsilon_C = 0.$$

Also, in this case the ESMC and REMC with the histogram method give essentially the same entropy approximations. Again, the only differences are in the region of the highest energies. The number of conformations in the lowest energy bin is large. This is a qualitatively different picture than that seen for *HP* chains, where the number of folded conformations is small, or as for S1 there is just one folded state (with accuracy to the mirror image).

### VI. CONCLUSIONS

In this paper a simple lattice model of globular proteins was studied. It was demonstrated that it is relatively easy to design *HP* sequences that fold into an (almost) unique globular state. The cooperativity of the folding transition results from a proper sequence design and from cooperative long-range interactions (explicit cooperativity of the hydrogen bond model). The higher cooperativity of the sequence with somewhat more complex *HP* patterns suggests that further diversification of the sequence patterns (by marking the turn regions by more *P*-type residues) may lead to more cooperative folding. The uniqueness of the folded state seems to be uncorrelated with the model's cooperativity.

The thermodynamics of the model proteins was investi-

gated by a combination of the replica exchange Monte Carlo method with the histogram method. This approach enabled a simultaneous estimation of the system energy and entropy. Comparison with the entropy sampling Monte Carlo method<sup>11-13</sup> (a version of the multicanonical ensemble method<sup>8,10</sup>) shows that the proposed procedure is accurate. At the same time, the REMC technique is computationally many times less expensive than simple METROPOLIS sampling or ESMC simulations. Since there is a relatively large conformational space accessible to the fcc chains, this suggests that the thermodynamics of more detailed models of proteins can be efficiently studied by means of the REMC method.

### ACKNOWLEDGMENTS

This work was supported by KBN Grant No. GP04A-1413 and by NIH Grant No. GM37408 of the Division of General Medical Sciences.

- <sup>1</sup>D. Gront, A. Kolinski, and J. Skolnick, *J. Chem. Phys.* **113**, 5065 (2000).
- <sup>2</sup>G. J. Geyer, *Stat. Sci.* **7**, 437 (1992).
- <sup>3</sup>K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- <sup>4</sup>R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- <sup>5</sup>U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- <sup>6</sup>N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **51**, 1087 (1953).
- <sup>7</sup>K. Binder, *Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Oxford University Press, New York, 1995).
- <sup>8</sup>B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1991).
- <sup>9</sup>J. Lee, *Phys. Rev. Lett.* **71**, 211 (1993).
- <sup>10</sup>U. H. E. Hansmann and Y. Okamoto, *J. Comput. Chem.* **14**, 1333 (1993).
- <sup>11</sup>M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 4940 (1994).
- <sup>12</sup>M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 9882 (1994).
- <sup>13</sup>M.-H. Hao and H. A. Scheraga, *J. Chem. Phys.* **102**, 1334 (1995).
- <sup>14</sup>A. Kolinski, W. Galazka, and J. Skolnick, *Proteins, J. Phys.* **26**, 271 (1996).
- <sup>15</sup>A. Kolinski, W. Galazka, and J. Skolnick, *J. Chem. Phys.* **108**, 2608 (1998).
- <sup>16</sup>D. Mohanty, A. Kolinski, and J. Skolnick, *Biophys. J.* **77**, 54 (1999).
- <sup>17</sup>U. H. E. Hansmann and Y. Okamoto, *Curr. Opin. Struct. Biol.* **9**, 177 (1999).
- <sup>18</sup>A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- <sup>19</sup>A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- <sup>20</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>21</sup>A. Kolinski and J. Skolnick, *Lattice Models of Protein Folding, Dynamics and Thermodynamics* (R. G. Landes, Austin, TX, 1996).
- <sup>22</sup>C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland, New York, 1991).
- <sup>23</sup>A. Kolinski, P. Rotkiewicz, B. Ilkowski, and J. Skolnick, *Prog. Theor. Phys. Suppl.* **138**, 292 (2000).