THERMODYNAMIC PROPERTIES OF POLYPEPTIDE CHAINS. PARALLEL TEMPERING MONTE CARLO SIMULATIONS*

Andrzej Sikorski†, Dominik Gront

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

(Received September 29, 2006)

A coarse-grained model of polypeptide chains was designed and studied. The chains consisted of united atoms located at the position of alpha carbons and the coordinates of these atoms were restricted to a [310] type lattice. Two kinds of amino acids residues were defined: hydrophilic and hydrophobic ones. The sequence of the residues was assumed to be characteristic for α-helical proteins (the helical septet). The force field used consisted of the long-range contact potential between residues and the local potential preferring conformational states, which were characteristic for α-helices. In order to study the thermodynamics of our model we employed the Multi-histogram method combined with the Parallel Tempering (the Replica Exchange) Monte Carlo sampling scheme. The optimal set of temperatures for the Parallel Tempering simulations was found by an iterative procedure. The influence of the temperature and the force field on the properties of coil-to-globule transition was studied. It was shown that this method can give more precise results when compared to Metropolis and Replica Exchange methods.

PACS numbers: 02.50.Ng, 02.70.Tt, 87.15.Aa

1. Introduction

In last decades one observed a rapid development of simulation methods applied to study the properties of polymer and biopolymer systems. Some interesting results were obtained, however, for most biopolymer systems, models were rather complicated and contained enormous amount of parameters. Therefore, it is sometimes difficult to judge which of them are really

* Presented at the XIX Marian Smoluchowski Symposium on Statistical Physics, Kraków, Poland, May 14–17, 2006.

† corresponding author.

important to obtain proper results. This was the main reason for designing and studying simple models of biopolymers. The most simplified models of protein folding were introduced some years ago by Dill and Shakhnovich and the main conclusion from these studies was that the compactness of the chain induced some amount of secondary structures [1,2]. However, recently Kolinski and Madziar [3] and Romiszowski and Sikorski [4, 5] pointed out that the presence of long-range interactions was not sufficient to generate the proper amount of secondary structures in the folded globule. A series of papers by Kuznetsov *et al.* was devoted to the analysis of the collapse transition with the respect to the hydrophobicitity of the chain [6, 7] while the effect of chain stiffness on the transition was studied by Frenkel *et al.* [8]. The effect of the chain length and stiffness on the coil-globule transition was studied recently by Binder *et al.* [9]. In the above-mentioned simple lattice models of Kolinski and Sikorski *et al.* [3, 10–12] polypeptide model chains were built on a lattice (310). The lattice used for the representation of proteins mimics the real structures with great accuracy [11]. In these models we introduced the primary sequence composed of two kinds of residues only (called hydrophobic and hydrophilic) interacting with specific potentials. It was shown that the formation of α-helices was marginal when no local preference of forming the helices was present. The interplay between long-distance tertiary interactions and a local helical potential during the folding process was also studied. The next problem that came up concerned the thermodynamic description of the coil-to-globule transition of polypeptide chains. For this purpose we used the Replica Exchange Monte Carlo method combined with the Histogram Method [13, 14]. By applying this method we determined the thermodynamics of the folding process and, moreover, the accuracy of the calculations appeared to be much more precise when compared to the classical Metropolis-like Monte Carlo method.

In this work we continue this approach to the protein folding problem. For this purpose we used the model mentioned above. We used a Replica Exchange Monte Carlo simulation algorithm combined with the Multi-histogram method. Such a combination allows one to obtain a very accurate thermodynamic description of a system of interest. In order to obtain the highest possible accuracy we applied a novel scheme for selecting a set of temperatures used in simulations.

2. The model

In our model we replaced a real polypeptide chain with a sequence of statistical segments connected by united atoms located on the positions of alpha carbons while the remaining atomic details were suppressed. Such a chain was restricted to a lattice of 310 type, which was frequently used in simulations of biopolymers [3–5]. In this lattice we can reproduce the

conformation of real proteins with the accuracy lower than 1 Å when compared our structures with the real ones when the lattice unit was equal to 1.22 Å. In order to make the model more realistic too small or too big angles between the consecutive segments were excluded. In our study we used the model chains consisting of two kinds of amino acid residues: hydrophobic (called H) and hydrophilic (called P). The sequence of residues in the chain was designed to mimic an idealised helical septet that can be found in real helical proteins. Therefore, the sequence — HPPHPP — was chosen following the idea of Hodges *et al.* [15]. The differentiation between these two kinds of residues was made by introducing the interaction potential of a pair of non-neighbouring residues. The potential V_{ij} had the following form:

$$V_{ij} = \begin{cases} \varepsilon_{\text{rep}} & \text{for } r_{ij} < r_1\,, \\ \varepsilon & \text{for } r_1 \le r_{ij} < r_2\,, \\ 0 & \text{for } r_{ij} \ge r_2\,, \end{cases} \tag{1}$$

where r_{ij} is a distance between a pair of residues i and j for $|i-j|>2$, $r_1 = 3$ (lattice units) and $r_2 = 5$ (lattice units). The repulsion for short distances was assumed to be the same for all pairs $\varepsilon_{\text{rep}} = 5k_{\text{B}}T$. This finite value of the repulsion was chosen in order to help the chain to rearrange its conformation in the dense collapsed state. The potential ε for intermediate distances was assumed to take the following values: for a pair HH the potential was $\varepsilon_{\text{HH}} = -2k_{\text{B}}T$, for a pair PP $\varepsilon_{\text{PP}} = -1k_{\text{B}}T$ and for a pair HP $\varepsilon_{\text{HP}} = 0$. The above assignment of potentials of interaction should help in forming a hydrophobic core inside the chain collapsed at lower temperatures. This set of interactions, although was not very realistic for proteins, was found to be better than a pure "hydrophobic" potential and it guaranteed the proper ground state of the system [16]. The preference of forming the α-helices was also introduced into the model. Actually, we introduced the preference of the formation of the right-handed helix only. The helical state is formed by three consecutive vectors and can be identified from a value of the following expression:

$$(r^*_{i-1,i+2})^2 = (\vec{v}_{i-1} + \vec{v}_i + \vec{v}_{i+1}) \cdot \text{sign}((\vec{v}_{i-1} \times \vec{v}_i) \cdot \vec{v}_i)\,, \tag{2}$$

where $\vec{v}_{i-1}, \vec{v}_i, \vec{v}_{i+1}$ are three consecutive vectors (segments) connecting i–1-th i-th and i+1-th residues. A right-handed α-helical state corresponds to the values of $(r^*_{i-1,i+2})^2$ between 9 and 25 (lattice units squared) [4,6]. The appearance of a right-handed helical conformation in the chain during the simulation process was associated with the energy loss equal to the ε_{loc}.

3. The simulation method

The simulations were carried out using a combination of a Monte Carlo Replica Exchange Method (REMC) [17], also known as Parallel Tempering or Metropolis-Coupled Chain [18] and the Multi-histogram method. In the REMC method we performed M simultaneous simulations of the same chain (replicas) but each at the different temperature T_i. Each replica was a subject to the Metropolis sampling algorithm, where changes of chain conformations were made as a series of the chain local micro-modifications. The set of elementary motions consisted of (Fig. 1): 2-segment motions, 3-segment motions, end-of-the-chain reorientations. At certain time intervals a pair of neighbouring replicas i and $i+1$ was selected at random and the attempt of their exchange was made with the following probability [19]:

$$\begin{aligned} P_{i,i+1} &= \min(1, \exp(-\Delta)) \,, \\ \Delta &= \left(\frac{1}{k_{\mathrm{B}} T_{i+1}} - \frac{1}{k_{\mathrm{B}} T_i} \right) (E_{i+1} - E_i) \,, \end{aligned} \tag{3}$$

where T_i and T_{i+1} are the temperatures of the i-th and $i+1$-th replica respectively while E_i and E_{i+1} are their total energies, k_{B} is the Boltzmann constant. The replica exchanges lead to a random walk in the temperature space and enable configurations to cross energy barriers. The system can move out of local minima. Therefore, this sampling scheme provides an enhanced sampling of low-energy structures. The REMC is known as one of the most efficient Monte Carlo based minimisation schemes [19, 20]. During a REMC simulation one can measure a desired property of a system separately for each temperature. Thus, for each replica temperature a canonical

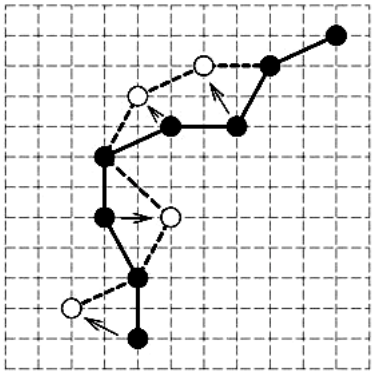

Fig. 1. A scheme of a polypeptide chain on a hybrid [310] lattice with the examples of the local changes of chain conformation.

average value of the property can be calculated. In order to compute the dependence of a given observable as a continuous function of the temperature one should apply the Multi-histogram technique [13], also known as WHAM (Weighted Histogram Analysis Method) [14]. In this method total energy distributions (histograms) were collected for each temperature during the simulation. From this data it is possible to determine the density of states $\Omega(E)$ for a given system.

The Multi-histogram procedure relies on a mutual overlap of the probability of states $P(E)$ in the neighbouring replicas. Therefore, the optimal temperature selection is an important factor influencing the accuracy of a thermodynamic study. In this work we established the optimal temperature set via an iterative procedure. We started from a set of uniformly spaced temperatures. After a simulation conduced as described above, we computed the first approximation of the density of states of our system. Having known $\Omega(E)$ we can easily compute $P_T(E)$:

$$P_T(E) = \frac{1}{Z(T)}\Omega(E)\exp\left(\frac{-E}{k_{\mathrm{B}}T}\right), \tag{4}$$

where $Z(T)$ is the partition function of a given system. This leads us to a formula that describes an overlap between two distributions of states calculated for two temperatures $\mathrm{ove}(T_i, T_j)$ (see Fig. 2):

$$\mathrm{ove}(T_i, T_j) = \int \min(P_{T_i}(E), P_{T_j}(E))dE\,. \tag{5}$$

Our goal is to find such a set of temperatures, that keeps $\mathrm{ove}(T_i, T_{i+1})$ values constant for each pair of adjacent replicas. This paradigm has two features that are very useful for our simulations: *(i)* the constant overlap ratio guarantees the convergence of Ferrenberg–Swendsen reweighting technique *(ii)* because $\mathrm{ove}(T_i, T_{i+1})$ value is right-proportional to the ratio of accepted replica swaps [21], we also keep the ratio of accepted replica exchanges on a constant level and achieve fast random walk in the temperature space making our sampling yet more efficient. Obtaining such an optimal set of temperatures is relatively easy: starting from the first temperature T_1 one should look for a temperature value T_2 that keeps $\mathrm{ove}(T_1, T_2)$ equal to an assumed value. Such calculation must be repeated for $\mathrm{ove}(T_2, T_3)$, ..., $\mathrm{ove}(T_n, T_{n+1})$ until certain criteria is fulfilled, *e.g.* the certain number of replicas were defined or the temperatures exceeded a certain value. Obviously, the number of temperatures in the set highly depends on the overlap value: the higher overlap, the greater number of temperatures. In this work we kept the overlap at the level of 0.85, while the number of replicas varied, from 35 to 55 for different values of $\varepsilon_{\mathrm{loc}}$ and different iterations. The whole procedure has been illustrated in Fig. 2.

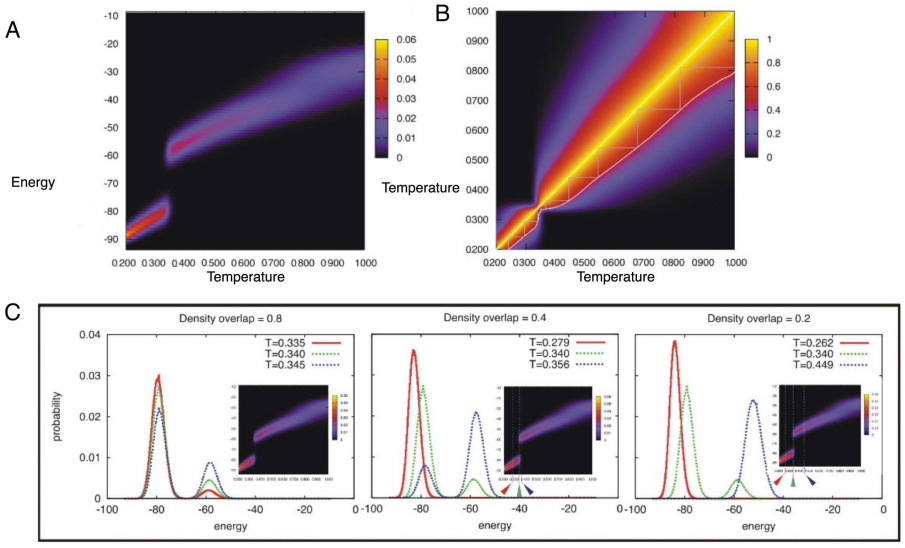

Fig. 2. A schematic overview of the algorithm for optimal temperature set selection. First, conditional probability distribution of energy $P(E|T)$ is computed for any temperature value from a predefined range (A). Then $P(E|T)$ dependence can be used to compute ove(T_i, T_j) function, that is an overlap ratio between two distributions: $P(E|T = T_i)$ and $P(E|T = T_j)$. The ove(T_i, T_j) is shown on (B). The optimal set of temperatures, discussed in this work, must keep ove(T_i, T_{i+1}) value constant for any pair of adjacent temperatures. Such a set can be found by walking on the isoline of ove(T_i, T_{i+1}) plot (narrow white line). Each plot in the panel (C) presents three distributions of states for the three temperatures: below, at and above the transition point. The three plots in the panel (C) differs in the overlap ratio (from the left: 0.8, 0.4 and 0.2).

4. Results and the discussion

The simulations were performed for linear chains consisted of $N = 60$ amino acid residues. The starting temperature range was from $T = 1$ to $T = 4$ basing on our previous findings in order to cover the random coil state of the polymer at high temperatures as well as the folded structures at low temperatures [5]. The local helical potential ε_{loc} took value 0 (flexible chains) and $-8k_{\text{B}}T$ (a very strong preference to form helical states).

The changes of the polymer chain's size during the annealing process are presented in Figs. 3(a), 3(b), where the mean-squared radius of gyration of the chain $\langle S^2 \rangle$ was calculated directly from simulations. One can observe that the size of the chain decreases smoothly with the temperature T.

This behavior corresponds to the transition from high-temperature random coil-like chains to densely packed globules at low temperature. The main difference between the flexible chain and the chain with the preference in forming secondary structures was the temperature of the transition: it was shifted towards higher temperatures for $\varepsilon_{\rm loc} = -8k_{\rm B}T$. One can observe that during the subsequent iterations of the Multi-histogram method the curve $\langle S^2 \rangle$ retains its shape but the position of the transition was shifted towards lower and higher temperatures. The number of iterations required to find the proper set of temperature and the $\langle S^2 \rangle$ curve was almost two times greater for the chain with non-zero local potential. This can be explained by the fact that in the latter case the energy hypersurface is more complicated.

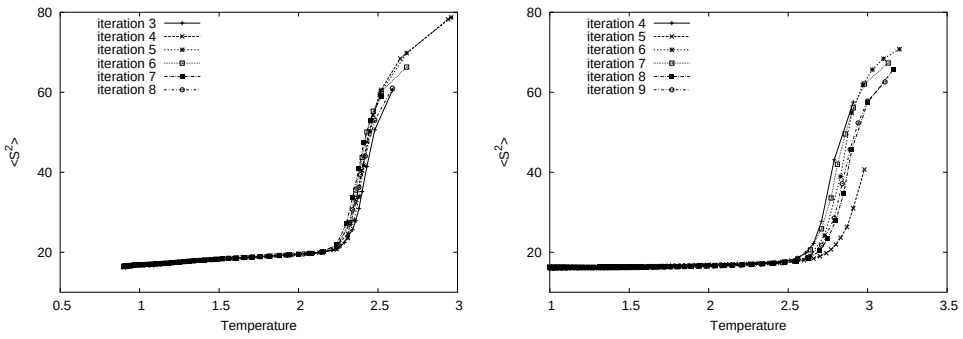

Fig. 3. The mean-square radius of gyration $\langle S^2 \rangle$ *versus* the temperature T for the chain with local potential $\varepsilon_{\rm loc} = 0$ (left plot) and $\varepsilon_{\rm loc} = -8k_{\rm B}T$ (right plot) obtained in subsequent iterations.

It is difficult to determine the temperature of the coil-to-globule transition from the behavior of the radius of gyration because this parameter does not change discontinuously. The estimation of the coil-to-globule temperature can be done from the behavior of the heat capacity. Fig. 4 presents an example of the heat capacity of the model chain $C_v/k_{\rm B}$ as a function of the temperature. The curves in this figure correspond to the same iterations of the Multi-histogram method as in Fig. 3(a). One can observe peaks on heat capacity curves located at temperatures corresponding to the rapid decrease of chain size. From the $C_v/k_{\rm B}$ plot one can estimate the coil-to-globule transition temperature: $T_{\rm C} = 2.385$ for $\varepsilon_{\rm loc} = 0$. It should be also noted that at very low temperature (T close to 1) another maximum on the heat capacity curve appears. This temperature cannot be related to any changes of chains size. It is difficult to explain the origin of this peak because it appeared at temperature where the simulation algorithm is rather ineffective: the fraction of the acceptance of local modifications of chain conformation

and the fraction of replica exchanges are rather low there. Probably some rearrangements (local ordering) in the dense collapsed chains takes place at this temperature.

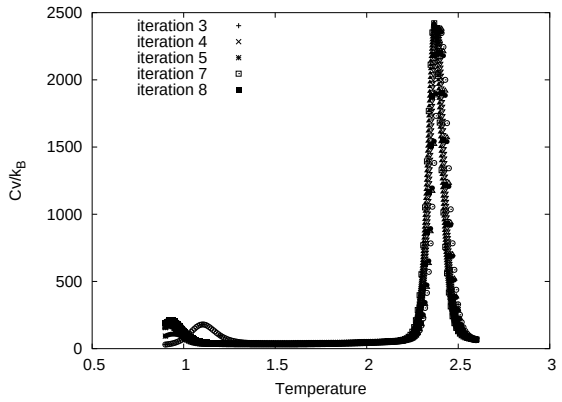

Fig. 4. The heat capacity C_v/k_B *versus* the temperature T for the chain with local potential $\varepsilon_\mathrm{loc} = 0$. The numbers of iterations are given in the inset.

It is possible to give the thermodynamic description of the system. Any observable $\mathcal{O}$ can be determined as a continuous function of temperature provided that an average value $\langle \mathcal{O} \rangle_E$ has been computed for each energy level E

$$\mathcal{O}(T) = \frac{1}{Z(T)} \sum_{E=E_\mathrm{min}}^{E_\mathrm{max}} \langle \mathcal{O} \rangle_E \, \Omega(E) \, \exp\left(\frac{-E}{k_\mathrm{B}T}\right) . \tag{6}$$

Fig. 5 presents an example of the free energy as a function of the total energy of the system. The presented plot was calculated for a temperature at which both minima on the curve corresponded to the same value of the free energy. Both states, high-temperature (a random coil) and low-temperature (a collapsed structure) have the same probability. Therefore, this case corresponds to the temperature of the coil-to-globule transition. The transition temperature from this chain can be thus estimated even more precisely than from a heat capacity plot: $T_\mathrm{C} = 2.862$. One can also observe that the barrier between these two states is rather small ($\sim 10 k_\mathrm{B}T$) and is symmetrical.

5. Conclusions

In this paper we studied the properties of simplified models of polypeptide chains. These chains were modelled as a linear sequence of united atoms that represented amino acid residues. The interplay between a tertiary potential and a local potential preferring helical states was studied by means

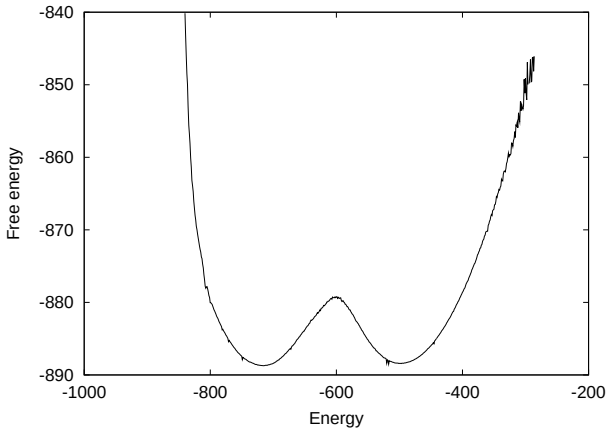

Fig. 5. The free energy F *versus* the total energy E. For the chain with the local potential the $\varepsilon_{\rm loc} = -8k_{\rm B}T$.

of Replica Exchange Monte Carlo simulations. The changes of the strength of the local potential did not change the size of collapsed chains significantly but shifted the coil-to-globule transition towards lower temperatures. The low-temperature conformations were not unique although they contain large fraction of secondary structures. The present work shows that the Replica Monte Carlo method combined with the Multi-histogram Method can be a useful tool for simulation of the simplified models of small proteins. It can give us the thermodynamic description of the system.

The computational part of this work was done using the computer cluster at the Computing Center of the Department of Chemistry, University of Warsaw.

REFERENCES

[1] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, H.S. Chan, *Protein Sci.* **4**, 561 (1995).

[2] K.A. Dill, *Protein Sci.* **8**, 1166 (1999).

[3] A. Kolinski, P. Madziar, *Biopolymers* **42**, 537 (1997).

[4] P. Romiszowski, A. Sikorski, *Physica A* **273**, 190 (1999).

[5] P. Romiszowski, A. Sikorski, *Biopolymers* **54** 262, (2000).

[6] Y.A. Kuznetsov, E.G. Timoshenko, *J. Chem. Phys.* **111**, 3744 (1999).

[7] Y.A. Kuznetsov, E.G. Timoshenko, K.A. Dawson, *J. Chem. Phys.* **103**, 4807 (1995).

[8] J.P.K. Doye, R.P. Sear, D. Frenkel, *J. Chem. Phys.* **108**, 2134 (1998).

[9] V.A. Ivanov, W. Paul, K. Binder, *J. Chem. Phys.* **109**, 5659 (1998).

[10] A. Kolinski, J. Skolnick, *Lattice Models of Protein Folding, Dynamics and Thermodynamics,* Landes, Austin, Texas 1996.

[11] A. Sikorski, J. Skolnick, *Proc. Natl. Acad. Sci. USA* **86**, 2668 (1989).

[12] A. Sikorski, J. Skolnick, *Biopolymers* **28**, 1097 (1989).

[13] A.M. Ferrenberg, R.H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).

[14] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, *J. Comput. Chem.* **13**, 1011 (1992).

[15] R.S. Hodges, A.K. Saund, P.C. Chong, R.E. Reid, *J. Biol. Chem.* **256**, 1214 (1981).

[16] E. Shakhnovich, A. Gutin, *Proc. Natl. Acad. Sci. USA* **90**, 7195 (1993).

[17] R.H. Swendsen, J-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).

[18] C.J. Geyer, in Proceedings of 23-rd Symposium on the Interface Foundation, Fairfax Station 1991.

[19] D. Gront, A. Kolinski, J. Skolnick, *J. Chem. Phys.* **113**, 5065 (2000).

[20] D. Gront, A. Kolinski, J. Skolnick, *J. Chem. Phys.* **115**, 1569 (2001).

[21] N. Rathore, M. Chopra, J.J. de Pablo, *J. Chem. Phys.* **122**, 024111 (2005).