Monatshefte für Chemie 137, 977–985 (2006)
DOI 10.1007/s00706-006-0485-1

**Monatshefte für Chemie**
**Chemical Monthly**
Printed in Austria

# Folding Behavior of Polypeptides. A Monte Carlo Study of Simplified Models

**Andrzej Sikorski*** and **Piotr Romiszowski**

Department of Chemistry, University of Warsaw, Warsaw, Poland

Received July 4, 2005; accepted January 23, 2006
Published online June 2, 2006 

**Summary.** A simple model of polypeptide chains was designed and studied. The chains were constructed on a flexible [310] lattice and consisted of united atoms located at the position of alpha carbons. Each united atom represented amino acid residues of two kinds: hydrophilic and hydrophobic. The sequence of the residues was assumed to be characteristic for $\alpha$- and $\beta$-type of proteins. The force field used consisted of the long-range contact potential between polymer segments, the short range repulsion, and the local potential preferring conformational states characteristic for $\alpha$-helices and $\beta$-strands. The Monte Carlo simulations of this model were carried out using the replica exchange technique coupled with the histogram method. The influence of temperature and the local potential on the size and internal structure of collapsed low temperature chains were studied. Thermodynamics of these systems consisting mainly of $\alpha$ and $\beta$ secondary structures were determined. The properties of the coil-to-globule transition were presented and compared with other theoretical predictions and simulation results.

**Keywords.** Coil-to-globule transition; $\alpha$ and $\beta$ globular proteins; Lattice models; Monte Carlo method; Replica exchange method.

## Introduction

The question of the folding transition in globular proteins is still far from understanding in spite of the enormous amount of experimental and theoretical work [1, 2]. The protein folding process is a very complicated phenomenon because of its internal complexity: heteropolymer chain molecules consisting of some thousands of atoms that can take enormous numbers of conformations but at certain conditions take only one (native) conformation. The theory of the coil-to-globule transition was developed by *Grosberg* and *Kuznetsov* [3] but even for much simpler systems like a homopolymer chain model it was shown that three different transitions can be found: gas-liquid, liquid-solid, and polymorphic solid-solid [4, 5]. The most simplified models of protein folding were introduced and developed some

* Corresponding author. E-mail: sikorski@chem.uw.edu.pl

years ago, mainly by *Dill* and *Shakhnovich* [6, 7]. In these models, the compactness of the chain at low temperatures induced the presence of secondary structures. Some theoretical considerations and computer simulations of stiff homopolymers showed that for certain stiffness the collapse transition became discontinued. The low temperature structures, in the latter case, appeared to be not only dense globules but also torus-like species [8–13]. The differentiation of the structural elements of the chain led to the increase of the number of transitions during the annealing process [14]. Monte Carlo simulations of lattice heteropolymer chains with a more complicated force field led to the fold, which were $\beta$-barrel type proteins [15].

In this work we continue a different approach to the protein folding problem [16–18]. For this purpose a simplified model of polypeptides was designed and developed. In this model we used a united atom representation of polypeptide chains. In order to mimic real polypeptides, we also differentiated these united atoms (that represent amino acid residues). To speed up the calculations we introduced the lattice approximation of our model chains. We used a replica exchange Monte Carlo simulation algorithm to determine properties of the model chains for a wide range of temperatures from a random coiled state to dense packed globules resembling globular proteins. Thus, within the frame of this model we can study the interplay between both potentials: a long-distance square-well contact potential and a local potential that prefers certain conformations characteristic for real proteins ($\alpha$-helical and $\beta$-type states).

## The Model and the Method

Each model chain under consideration consisted of $N = 100$ residues. Amino acid residues in our chains were represented by united atoms located at alpha carbon positions. The positions of the amino acid residues in the chain were restricted to vertices of a lattice of a [310] type, which was frequently used in simulations of systems containing polypeptides and proteins [17]. Chains on that lattice resemble conformations of real proteins and one can represent real structures with great accuracy – *RMSD* close to 0.7 Å where the lattice unit was equal to 1.22 Å. In our model the chain consisted of two kinds of residues only: hydrophilic (or polar – denoted as *P*) and hydrophobic (or non-polar, denoted as *H*). The distinction between these two kinds of residues was introduced by the interaction potential. This potential had the following form (Eq. (1)) where $r_{ij}$ is a distance between a pair of residues, $r_1 = 3^{1/2}$ (lattice units) and $r_2 = 5$ (lattice units).

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

We have assumed that the $\varepsilon_a$ potential took the following values for different types of contacts: $\varepsilon_{HP} = 0$, $\varepsilon_{HH} = -2kT$, and $\varepsilon_{PP} = -1kT$. The rigorous excluded volume effect was softened in our model and the repulsive potential was $\varepsilon_{rep} = 5kT$. This choice was done in order to help the chain to rearrange its conformation in a collapsed state. The other part of the force field was the local potential $\varepsilon_{loc}$, which characterized the tendency of the chain to form $\alpha$-helical or $\beta$-type structures. Let

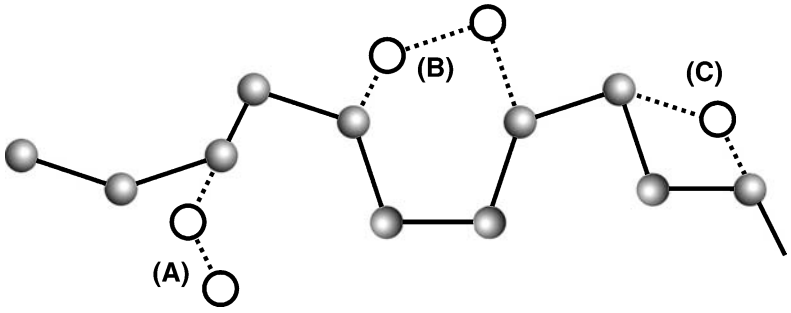


**Fig. 1.** A scheme of a polypeptide chain on a [310] lattice with examples of the two-bond motion (A), 3-bond motion (B), and 2-bond end reorientation (C)

consider the orientation of three consecutive vectors $\nu_{i-1}$, $\nu_i$, and $\nu_{i+1}$ using the following expression (Eq. (2)).

$$r^{*^2}_{i-1,i+2} = (\nu_{i-1} + \nu_i + \nu_{i+1})^2 \cdot sign((\nu_{i-1} \times \nu_i) \cdot \nu_{i+1}) \tag{2}$$

A right-handed $\alpha$-helical state corresponds to the values of $r^{*^2}_{i-1,i+2}$ located between 9 and 25 while $\beta$-type state corresponds to $-86 < r^{*^2}_{i-1,i+2} < -57$ and $57 < r^{*^2}_{i-1,i+2} < 91$ [16]. The occurrence of the right-handed $\alpha$-helical or $\beta$-type conformation in the chain during the simulation run was associated with the energy loss equal to the $\varepsilon_{loc}$.

The properties of model chains were determined by means of Monte Carlo simulations. We employed replica exchange Monte Carlo method (RMC) where the simultaneous simulations of the chains (replicas) are performed, each at the different temperature [19]. The distribution of temperatures is linear and each replica was a subject to the classical Metropolis sampling. The changes of the chain conformations were done using a set of local motions. This set consisted of: 2-bond motions, 3-bond motions, and 2-bond end reorientations (Fig. 1). During the simulation the replicas at neighboring temperatures were exchanged at random with the probability $P_{ij}$ (Eqs. (3) and (4)) [18, 19] where $T_i$ and $T_j$ are the temperatures of the $i$-th and $j$-th replica, respectively, while $E_i$ and $E_j$ are their total energies.

$$p_{ij} = \min(1, \exp(-\Delta)) \tag{3}$$

$$\Delta = \left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right) \cdot (E_j - E_i) \tag{4}$$

It was shown that the RMC simulations can be coupled with the histogram method in order to give the thermodynamic description of the system. In this method one collects histograms, showing the distribution of the total energy of the system.

Then, the entropy of the system can be determined using the definition of the function given by Eq. (5) where $S(E)$ is the entropy and $J(E)$ – the energy histogram [15, 18].

$$S(E)/k = \ln(J(E) + E/kT + constant \tag{5}$$

The other thermodynamic functions, like the free energy can be calculated by applying the relation of Eq. (6).

$$F(E,T) = E(T) - T \cdot S(E,T) \tag{6}$$

## Results and Discussion

The simulations were performed for chains consisted of $N = 100$ amino acid residues. In our investigations we studied pure $\alpha$ and $\beta$ proteins and, thus, our chains had amino acid residue sequences typical for both kinds of proteins. In the first case, it was an idealized helical septet -*HHPPHPP*-, which was repeated through the entire chain. For $\beta$ polypeptides a -*HP*- sequence was chosen and repeated in the entire chain as a characteristic $\beta$-type structure. The temperature was varied in the range from 1 to 4 based on 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. The local potential $\varepsilon_{loc}$ took values 0, $-1$, $-2$, $-4$, and $-8$ in both cases of polypeptide chains under consideration.

The folding transition (coil-to-globule) of a polypeptide chain can be identified by the analysis of the changes of polymer size with the temperature. The size of a chain is described by the mean-squared radius of gyration $\langle S^2 \rangle$. Figure 2 presents the changes of this parameter with the temperature for $\alpha$ and $\beta$ chains for some values of the local potential. In all cases under consideration all curves have the sigmoidal shape what suggests the presence of the coil-to-globule transition. The size of chains at high temperatures is considerably larger for $\beta$ case apparently because of the presence of larger number of expanded ($\beta$) conformations. The increase of the local potential changes the $\langle S^2 \rangle$ curves in the following way. First, it shifts the transition towards higher temperatures for $\alpha$ chains and towards lower temperatures for $\beta$ chains. The reason of this behavior will be discussed below. Second, the changes of the chain's size at high temperatures caused by the local potential are quite opposite to those of $\alpha$ and $\beta$ chains. The size of low temperature structures does not depend on the strength of the local potential but in general for $\beta$ chains it is more than twice larger than for $\alpha$ chains. One has to note that for $\beta$ chain with $\varepsilon_{loc} = -8$ its size increases when the temperature is lowered. In general, the transition is more rapid for $\alpha$-helical chains.

The heat capacity of a polypeptide chain was calculated as a variance of the mean total energy. In Fig. 3 we present the heat capacity $C_\nu/k$ as a function of the

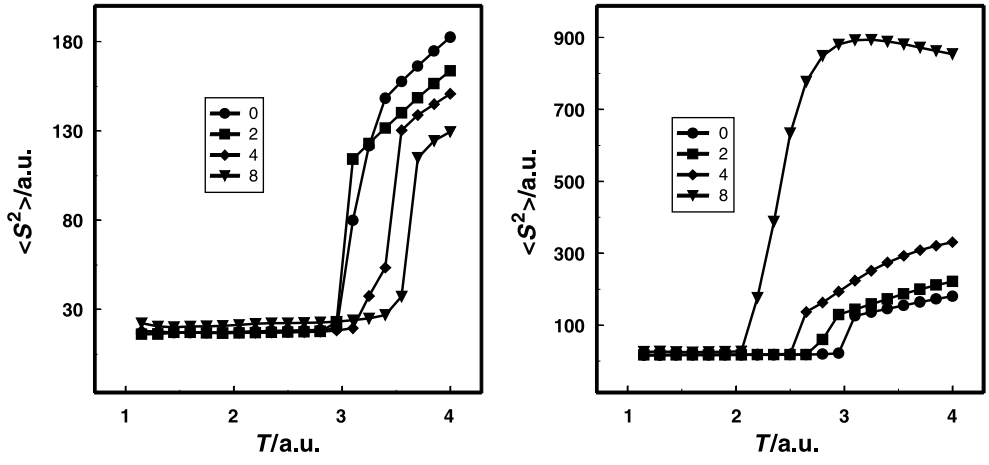


**Fig. 2.** The mean-square radius of gyration $\langle S^2 \rangle$ *versus* the temperature $T$ for the alpha (left) and beta (right) polypeptides; the strengths of the local potential are given in the inset

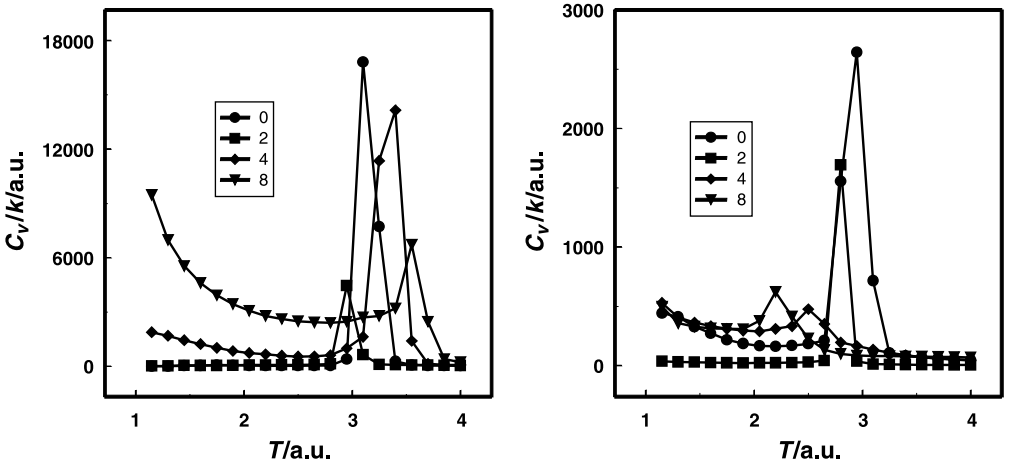


**Fig. 3.** The heat capacity $C_\nu/k$ *versus* the temperature $T$ for the alpha (left) and beta (right) polypeptides; the strengths of the local potential are given in the inset

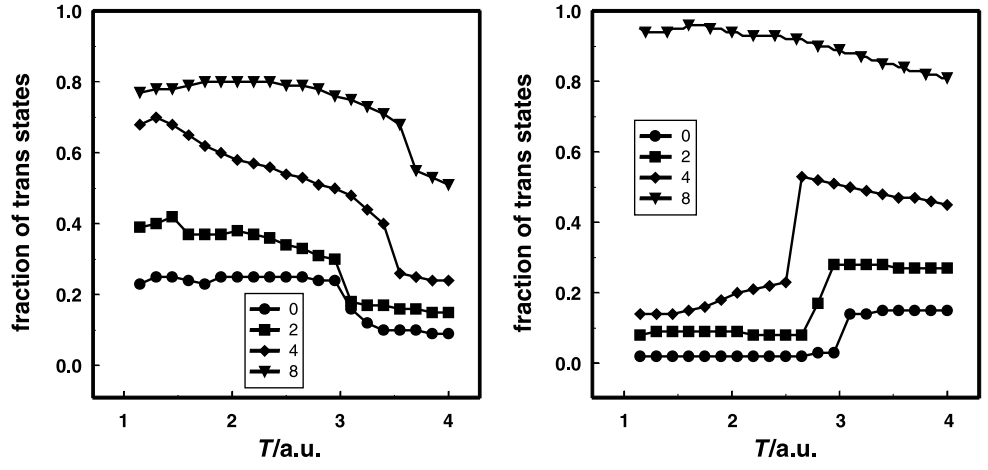


**Fig. 4.** The mean fraction of $\alpha$ and $\beta$ states *versus* the temperature $T$ for the alpha (left) and beta (right) polypeptides; the strengths of the local potential are given in the inset

temperature $T$ for various values of the local potential $\varepsilon_{loc}$. One can observe that the curves are rather smooth with one well-defined maximum. The positions of peaks on Fig. 4 are in full agreement with rapid decrease of the radius of gyration what confirms the presence of the folding transition. The values of $C_\nu/k$ at peaks for $\alpha$ chains are larger than for $\beta$ chains but their widths are smaller. This was apparently caused by the more rapid coil-to-globule transition for $\alpha$ chains.

In Fig. 4 one can observe the influence of the temperature on the mean fraction of $\alpha$-helical and expanded ($\beta$) states. At high temperatures chains contain *ca.* 12% of right-handed $\alpha$-helical states and *ca.* 15% of $\beta$ (expanded) states. In all cases of $\alpha$ chains under consideration these fractions (called helicity) increase during the annealing. One can observe a distinguished rapid increase of the helicity at the transition temperatures. It is interesting that even for $\varepsilon_{loc} = 0$ the helicity increased more than twice during the annealing. This means that the tertiary interactions plus

the proper helical-like sequence induced at low temperatures some helical states but the number of these states are much smaller than in real $\alpha$-helical proteins. Hence, the low temperature chains with stronger local potential have some features of $\alpha$-helical globular proteins. The introduction of the very strong local potential generates the helicity on the level 70–80%. The behavior of the fraction of expanded states in $\beta$ chains is different. For the local potential $\varepsilon_{loc} = 0$, $-1$, $-2$, and $-4$ the curves have similar shapes. Above the transition the fraction is almost constant or increases slightly. At the transition temperature the fraction of $\beta$ states decreases rapidly and the degree of this decrease is proportional to the strength of the local potential. Further annealing does not change the fraction significantly. This behavior suggests that during the transition the chain collapses to a small dense globule destroying most of the expanded states. For $\varepsilon_{loc} = -8$ the fraction of expanded states increases during the entire annealing process approaching the value 95%. This increase of $\beta$ states takes place in spite of the fact that the size of the chain remains small as for other values of the local potential. Thus, for the strong local potential the low-temperature chain has the features of a folded globular $\beta$ protein. Of course, one has to remember that all low temperature structures obtained for the model under consideration, for $\alpha$ as well as for $\beta$ chains, are not unique as in real proteins. Although in our model chains the number of $\alpha$ and $\beta$ states is rather high, no longer $\alpha$-helices or $\beta$-strands were found, similarly to the results obtained for other models with fully flexible heteropolymer chains [16, 20].

Figure 5 presents examples of the distributions of different energy levels, i.e. the energy histograms $J(E)$. The histograms shown were collected at the temperatures at the folding transition. For the transition temperature the population of the higher

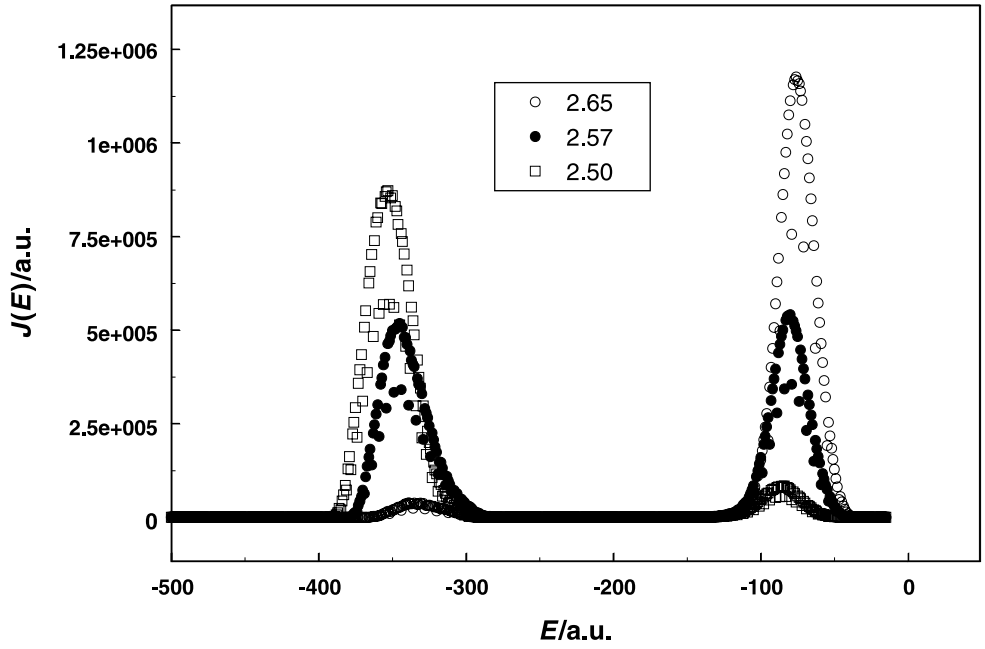


**Fig. 5.** The energy histogram $J(E)$ of the $\beta$ chain for the local potential $\varepsilon_{loc} = -1$; the temperatures are given in the inset

energy (unfolded) states is the same as that for lower energy (folded) states. For temperatures quite close to the transition (0.08 in our units) one can observe one peak only while the second one almost disappeared. At temperatures below the transition the peak corresponding to lower energies is higher than that correspond-

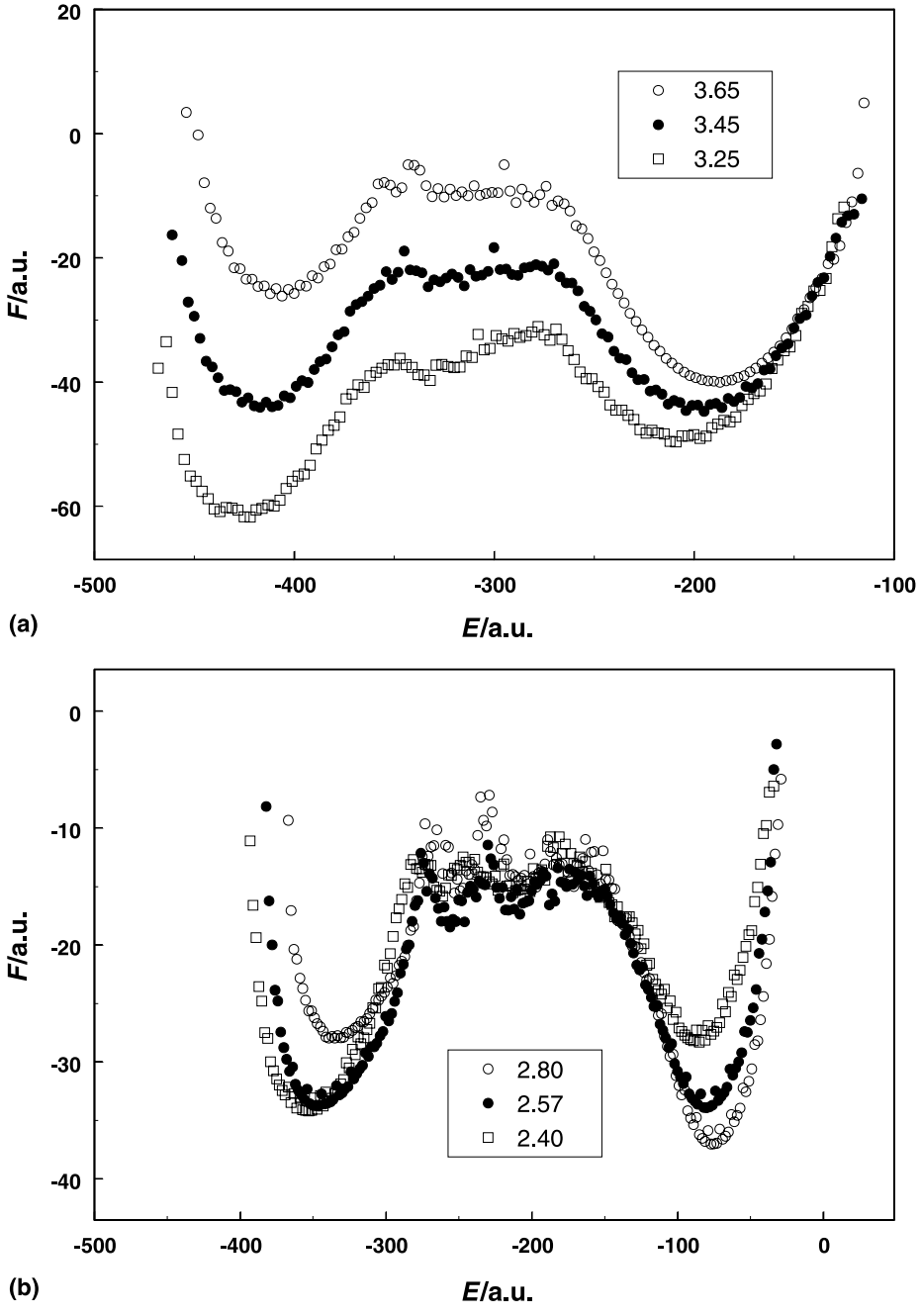


**Fig. 6.** The free energy $F$ *versus* the total energy $E$; the case of $\alpha$ chain with the local potential $\varepsilon_{loc} = -8$ (a) and of $\beta$ chain with the local potential $\varepsilon_{loc} = -1$ (b); the temperatures are given in the inset

ing to higher energies whereas at temperatures above the transition an opposite effect is observed. The comparison of the heights of the peaks can give the estimation of the transition temperature more precisely than $\langle S^2 \rangle$ or $C_\nu/k$ curves.

According to the Eqs. (5) and (6) we calculated the entropy $S$ and then the free energy $F$ of the system as a function of the total energy $E$. In Fig. 6 we present plots of the free energies as functions of the energy for some temperatures from the vicinity of the folding transition for $\alpha$ and $\beta$ chains respectively.

All free energy curves have almost the same shape with two minima corresponding to unfolded (on right) and folded (on left) states. Between these two states one can observe a free energy barrier. The free energy barriers between these two states of chain are similar for $\alpha$ chains and $\beta$ chains and changes between $15kT$ and $20kT$. The influence of the temperature on the location of a free energy curve is much more pronounced for $\alpha$ chains. One can observe that the influence of the temperature on the barriers is rather weak what suggests that the barriers are predominately of the entropic origin. The temperature also affects the location of the minima on the energy scale. One can also conclude that the higher value of the local potential increases the distance between free energy curves for different temperatures.

## Conclusions

In this paper we studied the properties of simplified models of polypeptide chains. These chains were modeled as a linear sequence of united atoms (beads) that represented aminoacid residues. The interplay between a tertiary potential and a local potential was studied by means of replica exchange Monte Carlo simulations. It was shown that in general the collapse transition differed for both sequences under consideration. 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. For the $\alpha$-helical polypeptides the introduction of the local potential leads to the formation of significantly higher numbers of helical structures during the folding process. For the $\beta$-type polypeptides a very strong local potential enforced similar behavior. The low-temperature conformations for both types of chains were not unique although they contain huge fraction of secondary structures. The histogram of energy obtained for one replica in the vicinity of the transition conditions served as a source of data for determining of thermodynamic characteristics of the system near the coil-to-globule transition.

## Acknowledgement

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] Pande V, Grosburg A, Tanaka T (2000) Rev Mod Phys **72**: 259
[2] Mirny L, Shakhnovich E (2001) Annu Rev Bioph Biom **30**: 361
[3] Grosberg AY, Kuznetsov DV (1992) Macromolecules **25**: 1970

[4] Zhou Y, Hall C, Karplus M (1996) Phys Rev Lett **77**: 2822
[5] Zhou Y, Karplus M, Wichert JM, Hall CK (1997) J Chem Phys **107**: 10691
[6] Dill K, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS (1995) Protein Sci **4**: 561
[7] Shakhnovich EI, Gutin AM (1993) Proc Natl Sci USA **32**: 7195
[8] Kuznetsov YA, Timoshenko EG, Davson KA (1996) J Chem Phys **105**: 7116
[9] Ivanov VA, Paul W, Binder K (1998) J Chem Phys **109**: 5659
[10] Doye JPK, Sear RP, Frenkel D (1998) J Chem Phys **108**: 2134
[11] Noguchi H, Yoshikawa K (1998) J Chem Phys **109**: 5070
[12] Kuznetsov YA, Timoshenko EG (1999) J Chem Phys **111**: 3744
[13] Sakaue T, Yoshikawa K (2002) J Chem Phys **117**: 6323
[14] Wang Y, Chen H, Liang H (2001) J Chem Phys **115**: 3951
[15] Gront D, Kolinski A, Skolnick J (2001) J Chem Phys **115**: 1569
[16] Kolinski A, Madziar P (1998) Biopolymers **42**: 537
[17] Romiszowski P, Sikorski A (2000) Biopolymers **54**: 262
[18] Sikorski A, Romiszowski P (2003) Biopolymers **69**: 391
[19] Smit B, Frenkel D (2002) Understanding Molecular Simulation, 2nd ed., Academic Press New York
[20] Magee JE, Warwicker J, Lue L (2004) J Chem Phys **120**: 11285

Verleger: Springer-Verlag GmbH, Sachsenplatz 4–6, 1201 Wien, Austria. – Herausgeber: Österreichische Akademie der Wissenschaften, Dr.-Ignaz-Seipel-Platz 2, 1010 Wien, und Gesellschaft Österreichischer Chemiker, Eschenbachgasse 9, 1010 Wien, Austria. – Redaktion: Getreidemarkt 9/163-OC, 1060 Wien, Austria. – Satz und Umbruch: Thomson Press Ltd., Chennai, India. – Offsetdruck: Manz Crossmedia, 1051 Wien, Austria. – Verlagsort: Wien. – Herstellungsort: Wien. – Printed in Austria.