Protein-like copolymers: computer simulation

P. G. Khalatur,[a] V. A. Ivanov,[b] N. P. Shusharina,[b] and A. R. Khokhlov[b]*

[a]*Department of Physical Chemistry, Tver' State University, 170002 Tver', Russian Federation. E-mail: khalatur@pcd.tversu.ac.ru*
[b]*Department of Physics, M. V. Lomonosov Moscow State University, Leninskie Gory, 117234 Moscow, Russian Federation. Fax: +7 (095) 939 2988. E-mail: shush@polly.phys.msu.su*

The model of *AB* copolymers with a "protein-like" primary sequence was developed. This type of copolymers was obtained in a computer experiment. First, the conformation of a collapsed dense homopolymer globule was generated and then, based on this conformation, the primary *AB* sequence was determined by denoting the monomeric units located near the surface of the globule as units *A* and those constituting the core of the globule as units *B*. After that, the primary structure of the chain was fixed, and different interaction potentials for the *A* and *B* units were introduced. Drawing an analogy of this model to aqueous solutions of globular proteins, *A* units were interpreted as hydrophilic, and *B* units were regarded as hydrophobic. By means of Monte Carlo simulation using the bond fluctuation model, the coil—globule transition in "protein-like" *AB* copolymer, induced by an increase in the attraction between the hydrophobic *B* units, was studied. The coil—globule transition in a copolymer with the "protein-like" primary sequence occurs at a higher temperature and has higher rate and is sharper than that in a random copolymer with the same *A/B* composition and in a random block copolymer with the same *A/B* composition and the same "degree of blockiness".

Key words: polymers, protein-like copolymers, coil—globule transition, computer simulation, Monte Carlo method.

Studies of microstructures in the systems of *AB* copolymers consisting of two types of monomer units (*A* and *B*) occupy a prominent place in polymer physics.[1—15] Block copolymers (with a block primary structure) and random copolymers (with a statistical primary structure) (see, for example, Refs. 1—7) are the *AB* copolymers studied most intensely. In some cases, studies on copolymers with short-range correlations along the chain are also made. These correlations always appear during copolymerization if the probability of addition of unit *A* or *B* to the growing chain at each particular step depends on the sort of unit added at the preceding step. The type of primary structure formed in this way can be characterized as *random with short-range correlations.*

In a rough approximation, globular proteins can be regarded as *AB* copolymers.[8—15] Indeed, the most important difference between the monomer units in globular proteins is that some amino acid residues are hydrophilic or charged, whereas the other residues are hydrophobic. Broadly, the index *A* can be ascribed to the former type of units, whereas the index *B* can be attributed to the latter type of units. Analysis of the primary *AB* sequence obtained in this way leads to the conclusion that the "protein-like" *AB* sequences are much more specific and provide more information than the simple primary structures described above. This is illustrated in Fig. 1. It is normally believed that in globular proteins, hydrophilic units *A* mostly cover the surface of globules and make the system stable to intermolecular aggregation, whereas hydrophobic units *B* form the core of the globule. Evidently, the requirement that in the dense globular state of an *AB* copolymer, the *A* units must be arranged on the surface, while the *B* units must be inside the core, is fairly strict, *i.e.*, it is satisfied for only a small fraction of all possible primary sequences. Moreover, since the *A/B* correlations thus introduced

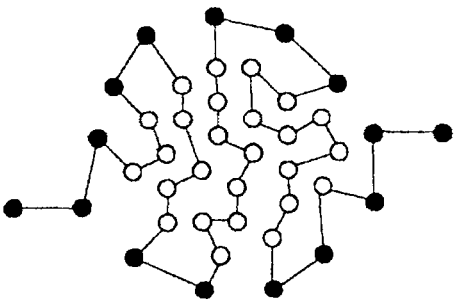

Fig. 1. Scheme of a globular protein. Hydrophobic units are denoted by light circles, while hydrophilic units are shown by dark circles.

Translated from *Izvestiya Akademii Nauk. Seriya Khimicheskaya*, No. 5, pp. 884—889, May, 1998.

1066-5285/98/4705-0855 $20.00

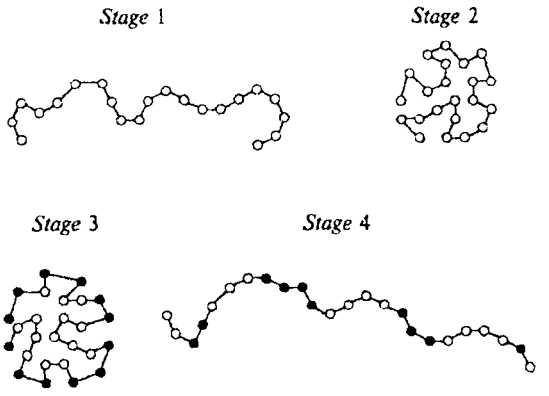

Fig. 2. Scheme of construction of the primary sequence of the "protein-like" *AB* copolymer.

depend on the conformation of the globule as a whole (*i.e.*, on the tertiary structure), they should be characterized as *long-range* correlations.

It is of interest to find out whether these primary structures can be obtained for *AB* copolymers (not necessarily of biological nature). This can be easily done in a computer simulation, whereas in a real experiment, this would be quite problematic. However, in both cases, the relevant procedure should include the following stages (Fig. 2).

Stage 1. A homopolymer coil swollen in a good solvent is considered.

Stage 2. Strong attraction between all the monomer units is "switched on", and a homopolymer globule is thus formed. In a real experiment, "switching on" the attraction means a sharp change in the temperature, addition of a poor solvent, *etc.*

Stage 3. This step is much more easily realized in a computer experiment. We take an "instantaneous picture" of the globule and "paint" the units located in the core and on the surface different colors, *i.e.*, the *A* index is attributed to those units that are located on the surface of the globule, and these units are called *hydrophilic*; index *B* is attributed to units situated in the core of the globule, and they are called *hydrophobic*. After that, this primary structure is fixed. In a real experiment, this "painting of the surface" can be accomplished using a reagent that reacts chemically with monomer units transforming hydrophobic units into hydrophilic or charged ones. If the quantity of this reagent is relatively small, one can expect that it would affect only those monomer units that are located on the surface of the globule, whereas its core would remain hydrophobic. Yet another important feature of this process is that the "painting" reaction should be relatively fast, while intermolecular aggregation, which always accompanies formation of globules, should be relatively slow. The aggregation can be retarded, for example, by using noncharged flow modifiers (thickening reagents).

Stage 4. This last stage is needed in the case of computer simulation. It includes "switching off" the uniform strong attraction between the monomer units and introduction of various interaction potentials for the *A* and *B* units.

Here we present the results of Monte Carlo (MC) computer simulation of the coil—globule transition for *AB* copolymers with "protein-like" primary sequences; this transition occurs upon an increase in the attraction between hydrophobic units *B* (for hydrophilic units *A*, the solvent always remains good, *i.e.*, there is no attraction between these units). The results obtained were compared with the corresponding data for random copolymers having the same *A*/*B* composition; the behavior of "protein-like" *AB* copolymers was found to differ substantially from that of random copolymers. Specific features of the "protein-like" primary *AB* sequence were also analyzed.

Copolymer Models and Calculation Procedure

In our computer experiment, an *N*-unit molecular chain consists of N_A monomer units of type *A* and N_B units of type *B* ($N = N_A + N_B$), located in the points of a simple cubic lattice. The molecules of a solvent (*S*) were represented by lattice vacancies. Standard bond fluctuation model was used.[9,10] In this model, it is assumed that each monomer unit of the chain occupies eight neighboring points of a simple cubic lattice, and the length of the bond vector between two units, neighboring along the chain, can fluctuate in the range from 2 to $\sqrt{10}$ steps of the lattice depending on the orientation of the bond vector. The allowed bond vectors ($\mathbf{b}$) form a set of 108 vectors, whose lengths $|\mathbf{b}|$ assume five different values: $|\mathbf{b}| \in \{2, \sqrt{5}, \sqrt{6}, 3, \sqrt{10}\}$. The repulsion between units of the same chain (interaction of the excluded volume) was modeled by prohibiting occupation of a single lattice point by two monomer units. As is usual, the MC procedure included stochastic selection of a monomer unit and an attempt to shift it in a random direction to one of the six neighboring lattice points.

Each configuration of the system is characterized by some energy of short-range interaction (U), which was defined in the following way. First, the condition of excluded volume requires that for any two monomer units that occupy the same point, $U = \infty$. Second, let $n_{\alpha\beta}$ be the total number of contacts between the corresponding units *A* and *B* that are the closest neighbors in the lattice, or between monomer units of the chain and solvent particles *S*. For example, n_{AS} means the total number of *A*—*S* contacts. Then

$$U = \sum_{\alpha\beta} \varepsilon_{\alpha\beta} n_{\alpha\beta}, \qquad (1)$$

where $\varepsilon_{\alpha\beta}$ are the corresponding parameters associated with the energy of interaction; for the sake of simplicity, below they are referred to as contact energies. It is necessary to mention that the energy parameters used here are related to Flory parameters ($\chi_{\alpha\beta}$) through the relations $\chi_{\alpha\beta} = z\varepsilon_{\alpha\beta}/(k_B T)$, where z is the effective coordination number of the lattice, T is the absolute temperature, and k_B is the Boltzmann's constant. (In what follows, it is assumed that temperature is expressed in energy units, and for simplicity k_B is assumed to be 1.) It is clear that among the short-range interaction parameters $\varepsilon_{\alpha\beta}$, the main parameters that determine the globular organization are ε_{AA}, ε_{BB}, ε_{AB}, ε_{AS}, and ε_{BS}. We set ε_{AA} to be equal to zero; in this case, the *A*—*A* interaction is reduced to intramolecular interaction of the excluded volume in an athermal system. This

condition corresponds to the case of a good solvent, fully compatible with the given polymer. It was also assumed that $\varepsilon_{AB} = 0$, *i.e.*, that the *A* and *B* units are incompatible for purely steric reasons. For a globule to be sufficiently stable and contain the *A* and *B* units mostly outside and inside the globular core, we should introduce the conditions: $\varepsilon_{AS} < 0$, $\varepsilon_{BS} > 0$, and $\varepsilon_{BB} < 0$. The parameters ε_{BB} and ε_{BS} describe the energy of "hydrophobic" interactions between the "nonpolar" *B* units of the chain and the "polar" solvent. Therefore, the energies $\varepsilon_{BB}n_{BB}$ and $\varepsilon_{BS}n_{BS}$ can be regarded as the contribution of "hydrophobic" interactions to the total energy. Similarly, the energy $\varepsilon_{AS}n_{AS}$ corresponds to the contribution of "hydrophilic" interactions. Thus, the total energy of the macromolecule can be expressed by the relation

$$U = \varepsilon_{AS}n_{AS} + \varepsilon_{BS}n_{BS} + \varepsilon_{BB}n_{BB}. \qquad (2)$$

Then, it is natural to assume that $|\varepsilon_{AS}| = |\varepsilon_{BS}| = |\varepsilon_{BB}|$, due to the similarity of the physical mechanisms of the corresponding interactions. In the calculations, it was assumed that $\varepsilon_{AS} = -1$, $\varepsilon_{BS} = 1$, and $\varepsilon_{BB} = -1$, and the temperature T (expressed in terms of energy units) was regarded as the main variable parameter of the system.

In the system described above, a test shift of a monomer unit can change the local energy, since in the general case, the neighbors of this unit vary. The probability to find some configuration of the system is proportional to $\exp[-U/(k_BT)]$. According to the Metropolis[18] algorithm, a test configuration that satisfies the condition of excluded volume and the limitations on the bond lengths is realized with the probability of transition $P(\mathbf{r} \rightarrow \mathbf{r}') = \min(1, \omega)$, where $\omega = \exp\{-[U(\mathbf{r}) - U(\mathbf{r}')]/(k_BT)\}$, and $\mathbf{r}$ and $\mathbf{r}'$ denote the corresponding configurations. Time (t) is expressed in the MC steps per particle; on the average, each particle executes one (successful or unsuccessful) test shift at each step.

Let us consider three models of an *AB*-copolymer chain.

1. *The copolymer with a "protein-like" primary sequence* is built according to the rules described in the introduction. The scheme of the construction of the sequence includes the following stages. In a homopolymer chain with volume interactions occurring in the conformation of a swollen coil, we introduce strong attraction between all the monomer units. In this case, we set $T = 1$. As a consequence, a dense homopolymer "parent" globule is formed. Then we choose $N_A = N/2$ monomer units, which have the maximum number of contacts with lattice vacancies (*i.e.*, with the solvent) and call these units hydrophilic units of type *A*. As a rule, these units are arranged near the surface of the "parent" globule formed at low temperature. The remaining $N_B = N/2$ units of the chain, which are usually arranged in the central body of the globule, are regarded as hydrophobic units *B*. Thus, we specify the primary structure of the heteropolymer chain. This primary structure is characterized by the average lengths of hydrophilic (L_A) and hydrophobic (L_B) blocks and by a peculiar distribution of the *A* and *B* units along the chain. The heteropolymeric globule obtained in this way is equilibrated during $(2-3) \cdot 10^6$ MC steps at a given temperature, and then average physical parameters are measured over the subsequent $\sim 4 \cdot 10^6$ MC steps. The above-described scheme for the construction of a sequence is repeated many times. The average length of the *B* unit is an important parameter of the primary sequence (below this parameter is designated by *L*).

2. *The random copolymer* is a chain with a simple stochastic sequence of the *A* and *B* units. In this case, by definition, we have $N_A = N_B = N/2$ and $L_A = L_B = 2$. As the initial configuration of the system, we take a heteropolymer globule equilibrated during $(2-3) \cdot 10^6$ MC steps at a given temperature.

3. *The random block copolymer* has a primary structure characterized by the Poisson distribution

$$f(x) = \exp(-\lambda) \cdot (\lambda^x/x!), \quad x = 0, 1, \ldots, \lambda > 0, \qquad (3)$$

where the average length (λ) of the continuous *B* sequence is equal to *L*. In this case, we pick only those copolymers for which $N_A = N_B = N/2$. As in the previous case, the initial state is a heteropolymer globule, and the properties of the system are analyzed over the last $\sim 4 \cdot 10^6$ MC steps after the system had been equilibrated at a given temperature.

All the results were obtained for chains consisting of $N = 256$ or 512 units. Correspondingly, for *AB* copolymers, $N_A = N_B = 128$ or $N_A = N_B = 256$.

Results and Discussion

Figure 3 shows typical distributions of monomeric units *A* and *B* along the chains of "protein-like," random, and random block copolymers. From a comparison of the primary structures of random and "protein-like" copolymers, it can be seen that the average lengths

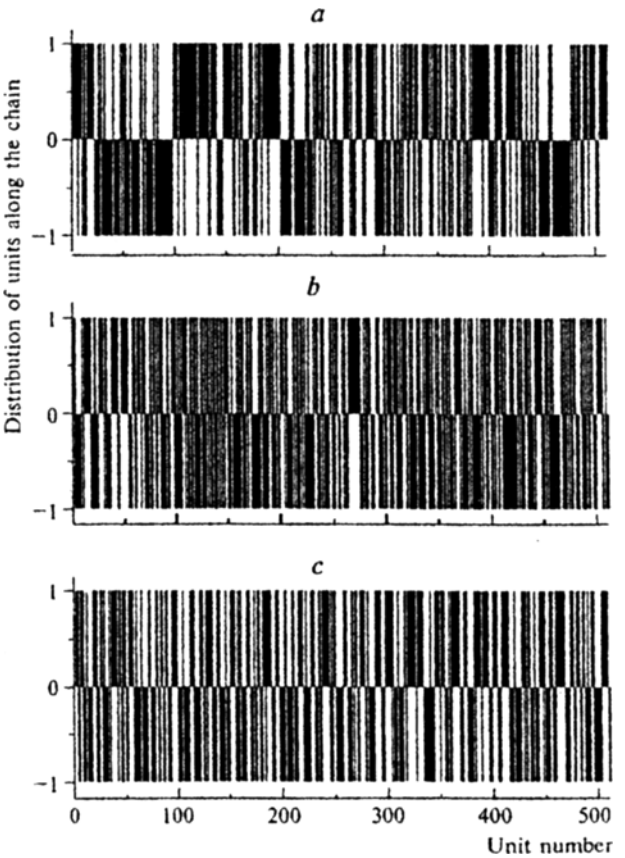

Fig. 3. Typical distributions of units *A* and *B* along the chain of a "protein-like" copolymer with $L = 3.173$ (*a*), random copolymer with $L = 1.984$ (*b*), and a random block copolymer with $L = 3.173$ (*c*). Units *A* are marked by −1, and *B* are marked by +1; $N = 512$.

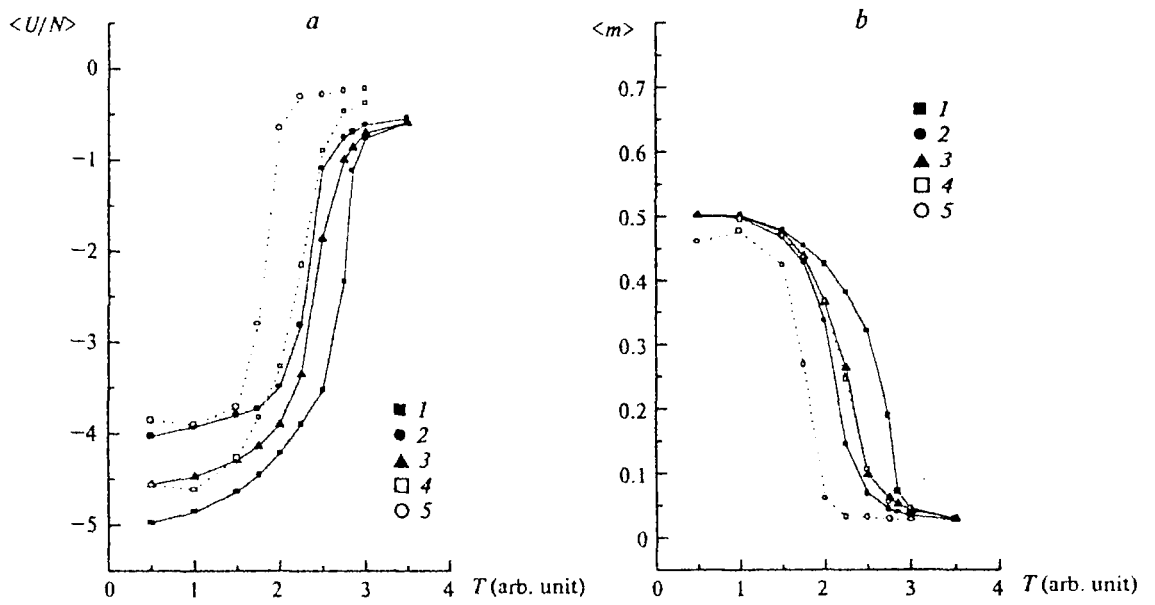

Fig. 4. Dependences of $<U/N>$ (*a*) and $<m>$ (*b*) on the temperature (T) for "protein-like" (L = 3.173), random (L = 1.984), and random block (L = 3.173) copolymers with various chain lengths: N = 512, "protein-like" (*1*); N = 512, random (*2*); N = 512, random block (*3*); N = 256, "protein-like" (*4*); N = 256, random (*5*).

of the A and B blocks in the "protein-like" copolymer are somewhat greater. However, it is obvious that in copolymers with a random block architecture (in which the average length of continuous sequences A and B is the same as in the "protein-like" copolymer), the distribution of the blocks along the chain is different. The main specific feature of the "protein-like" AB copolymers is the presence of relatively long sequences of the A and B monomer units. Thus, the primary structure of these copolymers can be characterized as *random with long-range correlations*.

Now we compare the characteristic features of the coil—globule transition in "protein-like" AB copolymers with those in random copolymers with the same A/B composition. We calculated the average potential energy per monomer unit ($<U/N>$) and the average aggregation number of the globular core ($<m>$). Note that the parameter $<m>$ is defined as the average number of hydrophobic units B forming the lattice-bound cluster. At equilibrium, the process of aggregation is characterized by the distribution function

$$W(m) = <N(m)>/\sum_m <N(m)>, \qquad (4)$$

where $<N(m)>$ is the (time) average number of m-particle aggregates built of monomer units B; thus, the average number of aggregation for units B is specified by the following formula:

$$<m> = \sum_m mW(m). \qquad (5)$$

Figure 4 presents the dependences of the $<U/N>$ and $<m>$ values on the temperature (expressed in arbitrary energy units) for the three types of copolymers under consideration. The temperature dependences of the derivatives $\partial<U/N>/\partial T$ and $\partial<m>/\partial T$, obtained by numerical differentiation of the corresponding plots shown in Fig. 4 for a chain consisting of 512 units, are presented in Fig. 5. Note that the derivative $\partial<U/N>/\partial T$ corresponds to the heat capacity per monomer unit and characterizes the fluctuations of the inner energy U at thermal equilibrium. For all copolymers, transition from the globular state (low temperatures) to the coil state (high temperatures) occurs within a relatively narrow temperature range in which the heat capacity and the $\partial<m>/\partial T$ value exhibit sharp peaks. Nevertheless, it can be seen that the coil—globule transition for the "protein-like" copolymer occurs at higher temperatures and is sharper than those for the random and random block copolymers with the same A/B compositions and the same L. Thus, it can be concluded that the specific primary structure inherited by "protein-like" copolymers from the "parent" globule shifts the coil—globule transition to higher temperatures and accounts for the fact that the resulting globule is more stable than those formed by random copolymers.

Now we discuss the effect of the total length of the chain on the behavior of the system. It can be easily seen from Fig. 4 that the behaviors of the chains with N = 256 and N = 512 qualitatively coincide. The quantitative differences are due to the fact that upon lengthening of the chain, the temperature of the coil—globule transition increases, whereas the average energy calculated per unit decreases. Both findings can be explained in the following way: the influence of the surface of the globule is more pronounced in the case of shorter chains, and this results in a different local average distribution

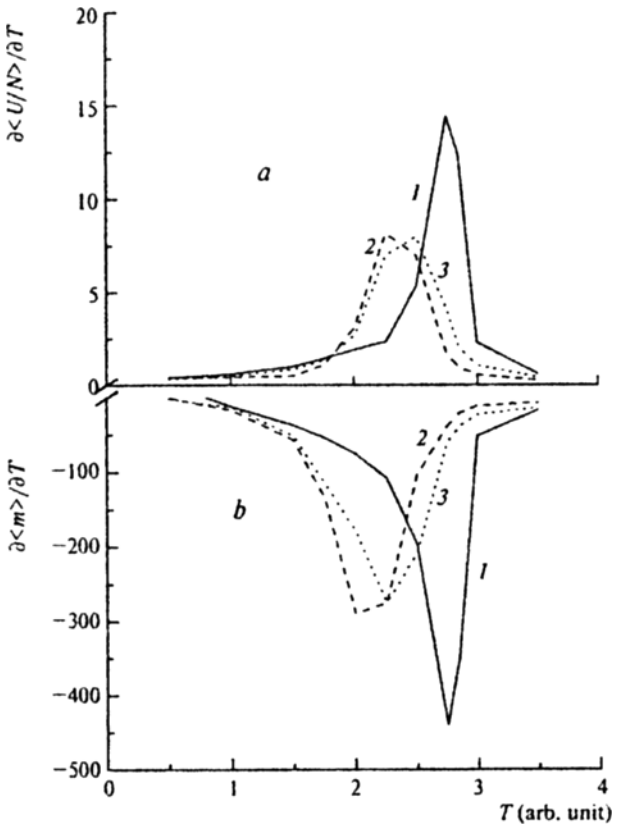

Fig. 5. Temperature dependences of $\partial\langle U/N\rangle/\partial T$ (*a*) and $\partial\langle m\rangle/\partial T$ (*b*) for 512-unit heteropolymer chains: "protein-like" (*1*); random (*2*); random block copolymer (*3*).

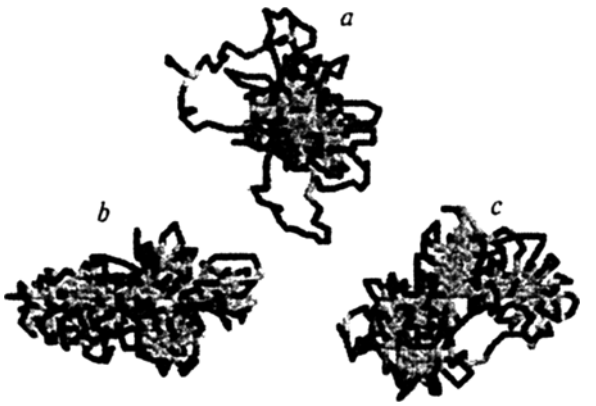

Fig. 6. Typical "instantaneous pictures" of the globular structures formed by "protein-like" (L = 3.173) (*a*), random (L = 1.984) (*b*), and random block copolymers (L = 3.173) (*c*) at T = 1.5 and N = 512. Light segments denote the hydrophobic B groups, and dark segments mark the hydrophilic A groups.

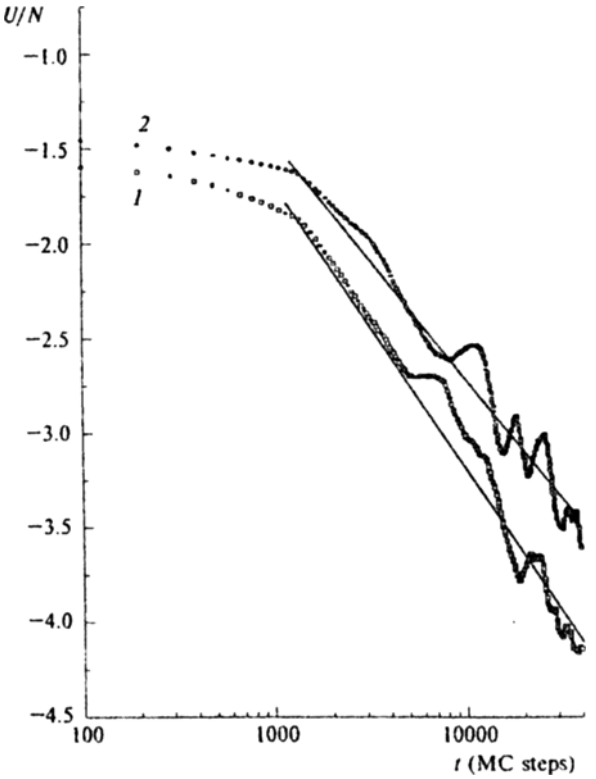

Fig. 7. Time evolution of U/N for 512-unit chains of "protein-like" (*1*) and random block (*2*) copolymers (L = 3.173) at T = 1.5.

of the units, which is looser. Therefore, the negative energy of interaction in a shorter chain is smaller in magnitude than that characterizing a longer chain, and the collapse of the shorter chain requires a more substantial cooling.

Let us consider the morphology of a heteropolymeric globule. Figure 6 presents typical "instantaneous pictures" of the globular structure obtained for the three types of copolymers at equilibrium at a low temperature (T = 1.5). It can be seen that the globules formed by "protein-like" AB copolymers possess a specific micellar morphology and contain a very dense core, consisting of hydrophobic groups B, and long loops of hydrophilic blocks A. Meanwhile, globules formed by random copolymers of both types contain larger, less dense, and less uniform cores and short surface loops.

We also studied the kinetics of the coil—globule transition. The coil resulting from thermal "denaturation" of the corresponding globules with a particular primary structure was chosen as the initial configuration in the simulation. Typical results of the simulation are presented in Fig. 7, which shows the plots for the time variation of the inner energy of "protein-like" and random block AB copolymers after the sharp change in the temperature to the globular region (T = 1.5). It can be seen that at early stages, the coil—globule transition in AB copolymers with "protein-like" primary sequences proceeds faster than that in random block copolymers.

* * *

Thus, a new model of an *AB* copolymer with a "protein-like" primary sequence was developed, and an appropriate scheme for constructing this sequence was proposed. Copolymers of this type can be obtained using the "instantaneous picture" of a dense homopolymeric globule assuming that the monomer units arranged near the surface of the globule are units of type *A*, whereas the globular core is formed by *B*-type units. After fixing this primary structure of the chain, different interaction potentials can be introduced for the *A* and *B* units. This is mostly done bearing in mind an analogy with aqueous solutions of globular proteins; hence, the *A* units are considered as hydrophilic blocks, while the *B* units are regarded as hydrophobic. The coil—globule transition in "protein-like" copolymers following an increase in the attraction between the hydrophobic *B* blocks was studied by the Monte Carlo method using bond fluctuation model. The results obtained for these copolymers were compared with those for random *AB* copolymers. An analysis of the primary structure of "protein-like" copolymers made it possible to conclude that the "degree of blockiness" of the *AB* sequence in these polymers is higher than that for random copolymers. For comparison, *AB* copolymers with the "random block" primary structure (and average length of continuous *A* and *B* sequences identical to that in "protein-like" copolymers) were also considered. It was found that the coil—globule transition for "protein-like" *AB* copolymers occurs at higher temperatures and has faster kinetics than that for random copolymers with the same *A*/*B* composition and for random block copolymers with the same *A*/*B* composition and the same "degree of blockiness." The globules of "protein-like" copolymers consist of micelle-like cores formed by the hydrophobic *B* units and stabilized by long loops of the hydrophilic *A* units. Thus, "protein-like" copolymers "inherit" some properties of the "parent" globule, which is manifested as the presence of long-range correlations in the primary structure.

References

1. T. Kotaka, H. Ohnuma, and H. Inagaki, in *Colloidal and Morphological Behavior of Block and Graft Copolymers*, Ed. G. E. Moran, Plenum Press, New York, 1971.
2. J. D. Dawkins, *Block Copolymers*, Eds. D. C. Allport and W. H. Jones, Wiley, New York, 1973.
3. J. Pouchy, A. Zivny, and A. Sikora, *J. Polym. Sci., A-2*, 1972, **10**, 151; *J. Polym. Sci., C*, 1972, **39**, 133.
4. S. F. Edwards, *J. Phys. A: Math. Nucl. Gen.*, 1974, **7**, 332.
5. T. Tanaka, T. Kotaka, and H. Inagaki, *Macromolecules*, 1976, **9**, 581.
6. T. M. Birshtein, A. M. Skvortsov, and A. A. Sariban, *Macromolecules*, 1976, **9**, 888.
7. J. Bendler, K. Solc, and W. Gobush, *Macromolecules*, 1977, **10**, 635.
8. M. Eigen and R. Winkler-Oswatitsch, *Steps Towards Life: a Perspective on Evolution*, Oxford University Press, Oxford—New York, 1992.
9. E. I. Shakhnovich and A. Gutin, *Biophys. Chem.*, 1989, **34**, 187.
10. E. I. Shakhnovich and A. Gutin, *Proc. Nat. Acad. Sci. USA*, 1993, **90**, 7195.
11. V. S. Pande, A. Yu. Grosberg, and T. Tanaka, *Proc. Nat. Acad. Sci. USA*, 1994, **91**, 12976; *J. Chem. Phys.*, 1994, **101**, 8246.
12. H. Li, A. Helling, C. Tang, and N. Vingreen, *Science*, 1996, **273**, 666.
13. E. I. Shakhnovich and A. Gutin, *J. Chem. Phys.*, 1990, **93**, 5967.
14. E. I. Shakhnovich, *Phys. Rev. Lett.*, 1994, **72**, 3907.
15. A. Yu. Grosberg, *Usp. Fiz. Nauk*, 1997, **167**, 129 [*Physics Uspekhi*, 1997, **167** (Engl. Transl.)].
16. I. Carmesin and K. Kremer, *Macromolecules*, 1988, **21**, 2819.
17. H.-P. Deutsch and K. Binder, *J. Chem. Phys.*, 1991, **94**, 2294.
18. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087.

Received October 30, 1997;
in revised form November 19, 1997