Conformation-Dependent Sequence Design (Engineering) of *AB* Copolymers

Alexei R. Khokhlov
Physics Department, Moscow State University, Moscow 117234, Russia

Pavel G. Khalatur
Department of Physical Chemistry, Tver State University, Tver 170002, Russia
(Received 3 June 1998)

A sequence design scheme to generate special primary structures in *AB* copolymers is proposed. In this scheme the type which is attributed to a given monomeric unit (*A* or *B*) depends on the spatial position of this unit in some "parent" homopolymer chain conformation. We consider two possible parent conformations: globular (giving rise to "proteinlike copolymers") and adsorbed (generating "adsorption-tuned copolymers"). By means of the Monte Carlo simulations it is shown that the *AB* copolymers with specially generated sequences have the physical properties different from those for random *AB* copolymers, i.e., they "inherit" some features of the parent conformation. [S0031-9007(99)09025-0]

PACS numbers: 61.41.+e

In this Letter, we propose a new approach to the design of specific primary sequences for the copolymers consisting of monomeric units of two types (*A* and *B*). This approach is based on the "coloring" in two "colors" monomeric units of a homopolymer taken in some well-defined conformation (globular conformation, conformation of adsorbed chain, etc.), depending on the spatial position of the unit in this "parent" conformation. Our computer simulations show that copolymers with *AB* sequences generated in this way acquire a number of special functional properties which distinguish them from the *AB* copolymers with random or block primary structures. In a sense, we can say that some functional features of the parent conformation are "memorized" or "inherited") and then manifested in other conditions. This special *AB*-sequence design (engineering of *AB* copolymers) can be achieved not only in computer simulations but in a chemical laboratory as well. Further studies in this direction may have an important impact both in the problem of obtaining *AB* copolymers with special functional properties and in the problem of the understanding of basic principles of molecular evolution at its early stages.

The general idea just described will be first illustrated taking as an example a globular conformation of a polymer chain.

Globular proteins enzymes functioning in living systems are the products of molecular evolution. Their primary structures involve 20 possible types of monomeric units (20 amino acid residues), therefore globular proteins are much more complicated objects than *AB* copolymers. However, the most essential distinction between different monomeric units of proteins is that some of these units are hydrophobic, while others are hydrophilic or charged [1]; thus in a very rough approximation it is possible to represent a globular protein as a kind of *AB* copolymer. The spatial (ternary) structure of such a copolymer in the native state would then normally correspond to the structure in which hydrophilic units (*A* type) cover the globular surface and prevent different globules from aggregation, while hydrophobic units (*B* type) constitute the dense globular core.

It should be emphasized that from the viewpoint of the possible *AB* primary sequences the requirement that in the dense globular conformation all the *B* units are in the core while all *A* units belong to the globular surface layer is rather restrictive. Only a minor fraction of all possible *AB* sequences has this property. For proteins the corresponding sequences were selected in the course of molecular evolution, because otherwise globular enzymes would precipitate and would be unable to perform their biological functions. The detailed analysis of the problem of sequence design in the application to protein molecules can be found elsewhere [1–7].

The problem which we address in the present Letter is somewhat different. Unlike Refs. [1–7] we are not focusing our study on the theoretical explanation of the properties of real proteins, as they emerged as a result of biological evolution. Our aim is to learn whether it is possible to make with *synthetic* copolymers a step along the same line as molecular evolution, i.e., to synthesize in the laboratory (or obtain in the computer experiment) a true *AB* copolymer (with *two* sorts of units) with the following "proteinlike" property: In the most dense globular conformation all the *A* units should be on the surface of the globule, while the *B* units should form the globular core. For the sake of brevity we will refer to such *AB* copolymers as "proteinlike copolymers" (although only one aspect of protein structure is taken into account). Such copolymers should have interesting physical properties, for example, they should not precipitate when the dense globular conformation is formed (this is not the case for ordinary macromolecules [8,9]).

In the present Letter, we describe the computer realization of this idea, although the main steps of

 0031-9007/99/82(17)/3456(4)$15.00

the real laboratory experiments should probably be the same.

(i) At first, we prepared a dense globular conformation of a homopolymer chain by switching on strong attraction between all monomeric units [Fig. 1(a)]. Of course, in real experiments by switching on the attraction we should understand the jump of temperature, addition of a poor solvent, etc.

(ii) Then, we have taken the "instant snapshot" of the globule and "colored" the units on the surface and in the core in different colors, i.e., we assigned the index A to those units that are on the surface of the globule and called these units hydrophilic, and assigned the index B to the units in the core of the globule and called these units hydrophobic. Then we fixed this primary structure [Fig. 1(b); the structure shown in this figure will be further referred to as *parent globule*]. In real experiments the "coloring of the surface" can be done with the chemical reagent entering in the reaction with monomeric units and converting them from hydrophobic to hydrophilic or charged. If the amount of this reagent is small enough, and the coloring reaction is fast, only surface monomeric units will be contaminated, the core remaining hydrophobic.

(iii) Finally, the last step was to remove a uniform strong attraction of monomeric units and to acquire different interaction potentials for A and B units [Fig. 1(c)].

We have performed the procedure described above for the chains of $N = 512$ units using for simulations the Monte Carlo (MC) method and the freely jointed bond fluctuation model [10]. This model, which is now widely used for MC simulations of polymer systems [11] combines the advantages of lattice and off-lattice approaches: Being lattice in nature, it allows the fluctuation of the bond length between successive polymer links in a certain interval.

One-half of the links most close to the center of mass of a homopolymer globule in parent conditions was selected and called hydrophobic (B-type units), while the remaining links were identified as hydrophilic (A-type units).

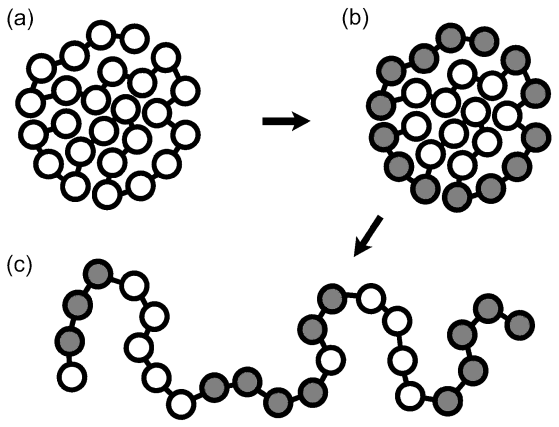

FIG. 1. Main steps of the sequence design scheme for proteinlike copolymers: (a) Homopolymer globule; (b) the same globule after the coloring procedure; (c) proteinlike copolymer in the coil state.

The typical primary sequence of proteinlike copolymer generated in this way had the "degree of blockiness" different from that of random AB copolymers: The average length of uninterrupted A or B sequences for proteinlike copolymers was equal to $\langle L\rangle = 3.173$, while for the random copolymer case it was $\langle L\rangle = 1.984$. This is not surprising, since if some monomeric unit lies at the globular surface there is an increased probability that the neighboring one also belongs to the surface layer.

We will see that many properties of proteinlike and random copolymers are different. In order to be able to distinguish whether this difference is due to the special sequence design explained in Fig. 1 or just due to the different degree of blockiness, we have introduced for comparison the "random-block" primary sequence. The random-block AB copolymers have the same average length L of uninterrupted A or B sequence as proteinlike copolymers ($\langle L\rangle = 3.173$), but in other respects the AB sequence is random; the distribution of the values of L was taken in the Poisson form, $f(L) = e^{-\lambda}\lambda^L/L!$ ($L = 0, 1, \ldots, \lambda$), where λ is equal to $\langle L\rangle$.

Let us compare the properties for the coil-globule transition for the AB copolymers with three different primary structures. The attractive interaction potential was introduced between the hydrophobic B units (with characteristic energy ε), while the interactions A-A and A-B were chosen to be purely repulsive (excluded volume interaction). Then, the coil-globule transition is realized upon temperature decrease, because of the increase of effective strength ε/k_BT of the attraction of B units.

We have obtained 5×10^3 different parent globules, then they have been colored and equilibrated during $(2-3) \times 10^6$ MC steps at a given temperature. The calculated properties for proteinlike AB copolymers (e.g., shown in Fig. 2) have been averaged over subsequent $\approx 4 \times 10^6$ MC steps and over the ensemble of 5×10^3 variants of coloring. The same type of averaging was

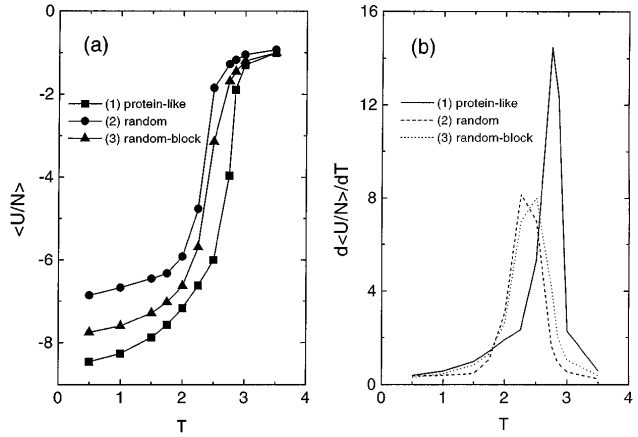

FIG. 2. The temperature dependencies (a) of the average energy of an attraction of B units in the globular state per one monomer link, $\langle U/N\rangle$, and (b) of the specific heat, $d\langle U/N\rangle/dT$, for (1) proteinlike, (2) random, and (3) random-block AB copolymers.

performed for the data for random and random-block copolymers. For more methodical details, see Ref. [12].

The Monte Carlo results for the temperature dependence of the average energy of attraction of hydrophobic B units per one monomeric unit, $\langle U/N \rangle$, are presented in Fig. 2(a), while the temperature derivative of the value of $\langle U/N \rangle$ (specific heat) is shown in Fig. 2(b). For all three types of primary sequences discussed above, it is possible to observe the change from the coil state (at high temperatures) to the globule state (at low temperatures) within the narrow temperature region where the specific heat exhibits quite a sharp peak. However, we can see that the coil-globule transition in the proteinlike copolymer occurs at higher temperatures and is much more abrupt (the peak in the specific heat is more pronounced), than for the random copolymer and random-block copolymer with the same A/B composition and the same value of $\langle L \rangle$ as for the proteinlike copolymer. The accuracy of all the data shown in Fig. 2 is within the interval from 5% to 10%, so the difference between the curves of Fig. 2 is quite meaningful.

We have also studied the kinetics of the coil-globule transition for the three types of the primary AB sequences discussed above. The general conclusion is that for proteinlike copolymers the process of the globule formation after the temperature jump is the fastest one. Also, the globule formed at the final stage is in this case most stable; this can be seen also from Fig. 2(a): The negative attraction energy of hydrophobic B units is the highest for proteinlike copolymers.

On the other hand, from Fig. 2 it is clear that although the results for random-block AB copolymers are between those for random and proteinlike copolymers, they are somewhat closer to the results for random AB sequences. Therefore, the main reason for the deviation of the properties of proteinlike copolymers from those for random copolymers is our special sequence design scheme (Fig. 1), not just a difference in the degree of blockiness.

This conclusion is further supported by the analysis of the morphology of the copolymer globules. Figure 3 shows typical instantaneous pictures (snapshots) of the globular structures obtained for three types of copolymers at equilibrium at low temperature ($T = 1.5$ in energetic units). It can be seen that for a proteinlike AB copolymer practically all B units are concentrated in the dense core of the globule which is stabilized by long dangling loops of hydrophilic A links. On the other hand, the core of the globules formed by random and random-block AB copolymers is much more loose and approximately 30% larger in average size than that for designed AB copolymers. Parts of hydrophilic A links are in this case inside the core, and those of them which belong to the surface form very short dangling loops which apparently are not sufficient to prevent the aggregation of such globules in the solution.

It is reasonable to assume that the formation of the dense core shown in Fig. 3 for a proteinlike copolymer

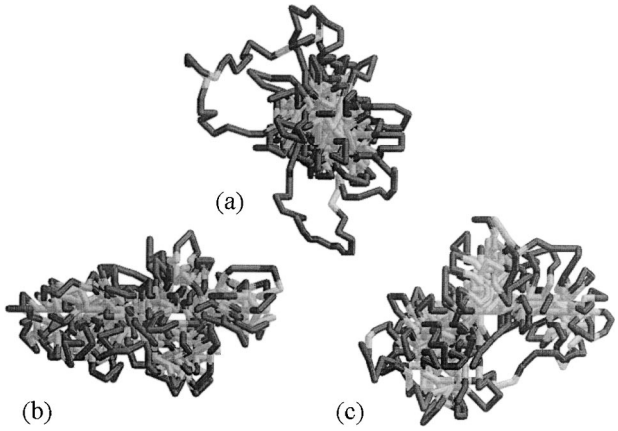

FIG. 3. Typical snapshots of the globular structures for (a) proteinlike, (b) random, and (c) random-block copolymers.

globule is facilitated by the fact that the dense globule preexisted in the parent conditions shown in Fig. 1(b). Since all the B links in this parent core are fitting to each other, there is no connectivity obstacles to *reassemble* most of this core when the effective attraction between the B units is switched on. In other words, we can say that the proteinlike copolymer inherited (or memorized) some important features of the parent globule which were then reproduced in the other conditions.

At this moment it is worthwhile to emphasize once more that a proteinlike copolymer globule cannot be regarded as a model of real globular proteins in the native state: The B core is liquidlike (although compact), while the spatial structure of a native protein is unique. Our ultimate target is a synthetic AB copolymer with useful functional properties (e.g., solubility of globules), rather than biological protein. In this sense we are interested here in "functional," rather than "structural" memory.

However, it is worth mentioning here some parallels with protein problems as well. The sequence design procedure of Shakhnovich and Gutin [4] as well as the imprinting method by Pande, Grosberg, and Tanaka [5] are similar in spirit to our approach, although these papers deal with different problems of uniqueness and stability of the native conformation of protein and its active catalytic center. Moreover, the method of Ref. [5] is connected with the threading of "monomer soup" equilibrated around the active center, which is different from the coloring procedure proposed above. In polymer language this corresponds to the difference between polymerization and chemical modification of a given polymer chain which can be a much softer chemical reaction. In Ref. [3] it was found that the number of minimum energy compact structures for random heteropolymers is increasing much more slowly than the total number of compact structures (or not increasing at all under certain conditions). This fact may be related to the sharpening of the coil-globule transition which we observed for proteinlike copolymers.

Let us now generalize the idea which was introduced above. Actually, the primary structure of a proteinlike

copolymer was generated via a kind of "coloring procedure" (in two colors, *A* and *B*) for a homopolymer chain in the globular state. However, the special primary sequence can be obtained not only from globular conformation; *any specific polymer chain conformation can play the role of a parent one.*

The simplest example of this kind is connected with the conformation of a homopolymer chain adsorbed on the plane surface. Let us color the links in direct contact with the surface in some typical instant snapshot conformation. This corresponds to the assumption that the surface catalyses some chemical transformation of the adsorbed links. Then we will end up with an *AB* copolymer for which the sequence design was performed in the parent adsorbed state. After desorption such an *AB* copolymer will have special functional properties: It will be "tuned to adsorption."

Indeed, we have performed Monte Carlo computer experiments along the lines of the sequence design scheme outlined above (see also Ref. [13]) for the chain of 32 monomer units. In the conformations of the adsorbed homopolymer chain eight units which are closest to the surface were identified and denoted as *B* units; others were designated as *A* units. Then we studied the adsorption behavior of the *AB*-copolymer chain obtained in this way on the plane surface induced by the increase of the attraction of *B* units to this surface and compared it with the behavior of the corresponding random and random-block copolymers (for details, see Ref. [13]).

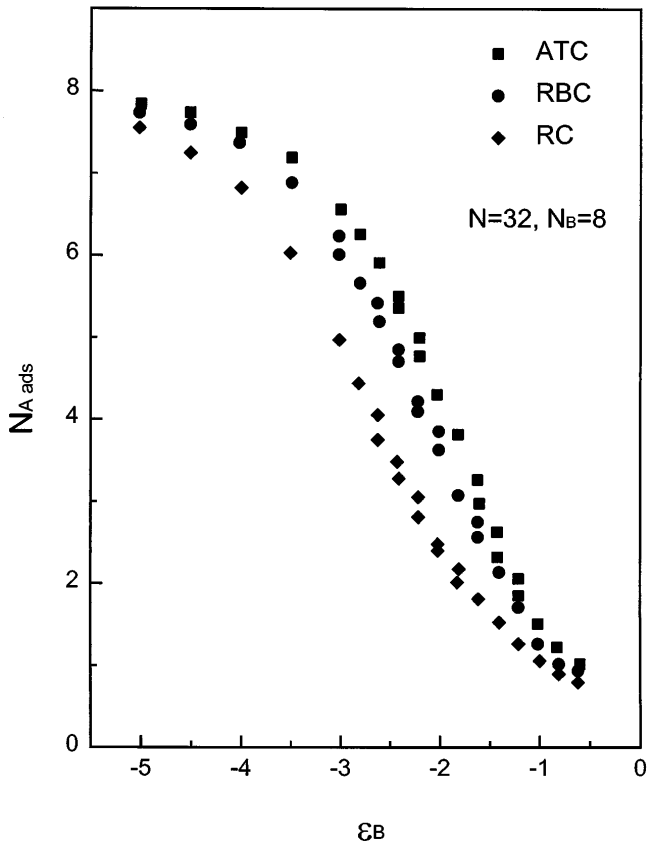

FIG. 4. The average number of adsorbed type-*B* segments vs the attraction energy to the surface, ε_B, for the adsorption-tuned (ATC), random-block (RBC), and random (RC) copolymer chains of length $N = 32$ with the number of type-*B* segments $N_B = 8$.

In Fig. 4 the plot of the average number of adsorbed *B* units vs the energy of their attraction to the surface, ε_B, is presented. It can be seen that the number of adsorbed segments (at a given value of ε_B) is always the highest for the designed *AB* copolymers. In other words, due to the memorizing of some functional features of the parent conformation we have indeed obtained the *AB* copolymer "tuned to adsorption on a plane surface."

In both examples considered above we generate the *AB* sequence with the help of some special *parent* conformation of a polymer chain. Such a conformation-dependent sequence design can be called "engineering of *AB* copolymers" using some remote analogy with protein engineering which deals with the design of the primary sequence of protein molecules.

It is important also to emphasize that in both cases some functional features of the parent conformation were memorized by the *AB* copolymers generated according to our sequence design scheme. These features are then manifested in other conditions. Such an interrelation can be regarded as one of the possible mechanisms of molecular evolution: polymer acquires some special primary sequence in the parent conditions and then (in other conditions) uses the fact that the primary structure is "tuned to perform certain functions."

[1] A. Yu. Grosberg, Phys.-Usp. **40**, 109 (1997).
[2] E. I. Shakhnovich, Folding & Design **3**, 45 (1998).
[3] C. J. Camacho and D. Thirumalai, Phys. Rev. Lett. **71**, 2505 (1993).
[4] E. I. Shakhnovich and A. M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1993).
[5] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Macromolecules **28**, 2218 (1995).
[6] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1996).
[7] E. D. Nelson, L. F. Teneyck, and J. N. Onuchic, Phys. Rev. Lett. **79**, 3534 (1997).
[8] I. M. Lifshitz, A. Yu. Grosberg, and A. R. Khokhlov, Rev. Mod. Phys. **50**, 683 (1978).
[9] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (American Institute of Physics, New York, 1994).
[10] I. Carmesin and K. Kremer, Macromolecules **21**, 2819 (1988).
[11] K. Kremer, in *Computer Simulation in Chemical Physics*, edited by M. P. Allen and D. J. Tildesley (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993), p. 397.
[12] A. R. Khokhlov and P. G. Khalatur, Physica (Amsterdam) **249A**, 253 (1998).
[13] E. A. Zheligovskaya, P. G. Khalatur, and A. R. Khokhlov, Phys. Rev. E **59**, 3071 (1999).