Monte Carlo simulation of AB-copolymers with saturating bonds

2003 J. Phys.: Condens. Matter 15 3013

(http://iopscience.iop.org/0953-8984/15/19/305)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **15** (2003) 3013–3027 PII: S0953-8984(03)56926-0

Monte Carlo simulation of AB-copolymers with saturating bonds

A V Chertovich[1], V A Ivanov[1], A R Khokhlov[1,2] and Jakob Bohr[3]

[1] Physics Department, Moscow State University, 119992 Moscow, Russia
[2] Department of Polymer Science, University of Ulm, 89069 Ulm, Germany
[3] Department of Physics, The Technical University of Denmark, DK-2800 Lyngby, Denmark

E-mail: chertov@polly.phys.msu.ru

Received 28 November 2002, in final form 13 February 2003
Published 6 May 2003
Online at stacks.iop.org/JPhysCM/15/3013

Abstract

Structural transitions in a single AB-copolymer chain where saturating bonds can be formed between A- and B-units are studied by means of Monte Carlo computer simulations using the bond fluctuation model. Three transitions are found, coil–globule, coil–hairpin and globule–hairpin, depending on the nature of a particular AB-sequence: statistical random sequence, diblock sequence and 'random–complementary' sequence (one-half of such an AB-sequence is random with Bernoulli statistics while the other half is complementary to the first one). The properties of random–complementary sequences are closer to those of diblock sequences than to the properties of random sequences. The model (although quite rough) is expected to represent some basic features of real RNA molecules, i.e. the formation of secondary structure of RNA due to hydrogen bonding of corresponding bases and stacking interactions of the base pairs in helixes. We introduce the notation of RNA-like copolymers and discuss in what sense the sequences studied here can be considered as RNA-like copolymers.

1. Introduction

To derive a better understanding of the physical principles behind the biological functioning of biopolymers (e.g. proteins, DNA, RNA) is one of the major challenges for science in the early part of the twenty-first century. It will allow an improved insight into the molecular principles behind the evolution of biopolymers and their sequences. An interdisciplinary scientific field is emerging; the study of simple models of these complex molecules is an important task. Some examples of recent work are the theory and computer simulations of different models for protein folding [1–12], conformation-dependent sequence design [13–18], the study of toroidal structure of DNA [19–24] and the secondary structure of RNA [25, 26].

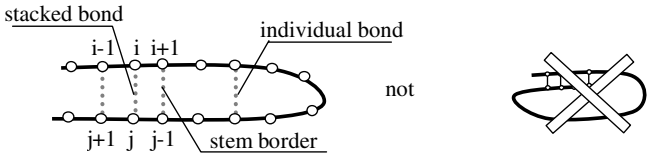


Figure 1. Allowed direction of strands forming a stem.

In this paper, we will consider a single AB-copolymer chain where the formation of saturating bonds (SBs) is allowed between A and B monomer units. The analysis of conformations of such a chain will be performed by means of Monte Carlo (MC) computer simulation. Three different kinds of primary AB-sequences are investigated: statistically random sequence, diblock sequence and 'random–complementary' sequence (one-half of such an AB-sequence is random with Bernoulli statistics while the other half is complementary to the first one). Each monomer unit of type A or B possesses one valency to form an SB with some monomer unit of opposite type. An important goal for our study is to understand how the SB formation influences the state diagram of possible spatial conformations of the copolymer chain. We are interested in finding the regions of stability of different conformations and in the location of transition lines between them. Subsequently, it will be possible to modify the primary sequence in the course of a simulation in order to study the effects of sequence evolution on the number of available conformations.

We keep in mind the formation of the secondary structure of real RNA molecules which is organized in the form of a 'clover leaf' as an example of the real features of biopolymers which serve as a guide for us in constructing our model. The present study lies within the framework of the recently proposed scheme of *conformation-dependent sequence design* [13–18].

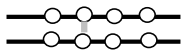
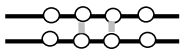
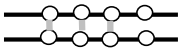
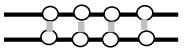
The paper is organized as follows. In section 2 we describe our model and discuss the importance of each of the assumptions made. In section 3 we present and discuss our results. Finally, section 4 contains conclusions and the outlook for further investigations.

2. Method: model and simulation technique

We use the bond fluctuation model in our MC simulation [27–29]. The molecule is a single N-unit chain with N_A monomer units of type A and N_B monomer units of type B ($N = N_A + N_B$). Each monomer unit occupies eight neighbouring lattice sites on a simple cubic lattice and the excluded volume condition is applied. The bond length between two successive monomer units along the chain can fluctuate between 2 and $\sqrt{10}$ lattice units.

The chain conformation is characterized by the interaction energy, E, between monomer units, which is defined in the following way. If a monomer unit of type A is in contact with a monomer unit of type B they can form a thermoreversible SB provided they are not successive along the chain. After one SB between two monomer units has been formed these units can no longer be involved in the formation of any other SB. Every SB adds $\varepsilon_{SB} < 0$ to the energy E of the system. Two (or more) SBs (pairs) formed between successive monomer units respectively along the chain are called a stem of SBs (i.e. the second pair should contain monomer units which are neighbours along the chain for the units of the first pair, see figure 1). Every border of the stem adds the positive value $-\varepsilon_{SB}\alpha_{coop}$ to the energy of the system, where we will call the parameter α_{coop} the factor of cooperativity. Introducing such a potential means that the borders are energetically unfavourable (see table 1 for an example of the SB-potential used).

Table 1. Dependence of energy on cooperativity parameter α_{coop}.

	α_{coop} =	1	0.5	0.0	General formulae
	E =	$-\varepsilon_{SB}$	0	ε_{SB}	$E = \varepsilon_{SB} \cdot (1-2\alpha_{coop})$
	E =	0	ε_{SB}	$2\varepsilon_{SB}$	$E = \varepsilon_{SB} + \varepsilon_{SB} \cdot (1-2\alpha_{coop})$
	E =	ε_{SB}	$2\varepsilon_{SB}$	$3\varepsilon_{SB}$	$E = 2\varepsilon_{SB} + \varepsilon_{SB} \cdot (1-2\alpha_{coop})$
	E =	$2\varepsilon_{SB}$	$3\varepsilon_{SB}$	$4\varepsilon_{SB}$	$E = 3\varepsilon_{SB} + \varepsilon_{SB} \cdot (1-2\alpha_{coop})$

Another possibility to account for this effect would be to introduce the energy gain for every stacking pair of SBs, i.e. the negative stacking energy. We do not consider this possibility here and do not discuss the equivalence of both approaches, although such a modification of the model is even more directly connected with the real physical processes of helix formation in biopolymers.

In our model it is specified that stems can be formed only in 'antiparallel mode', i.e. two strands forming a stem are always oppositely directed (see figure 1). That feature mimics some properties of real RNA molecules which are known to be directed polymers and to have the so-called 'head–tail' primary structure. Such a structure ensures the formation of helixes in real RNAs by only antiparallel strands. In our model the cooperativity bonus is only applied for antiparallel strands and not for parallel ones. Hence, the direction of strands which are forming stems does not play any role in the case $\alpha_{coop} = 0$.

In our simulation we used the following single MC steps:

(1) we choose a monomer unit randomly;
(2) we move that unit randomly to one of the six neighbouring lattice sites (this trial move is rejected if it breaks the excluded volume condition, i.e. if the monomer unit hits some already occupied lattice sites);
(3) if this monomer unit is involved in the formation of an SB we destroy this bond temporarily;
(4) we prepare the list of all neighbours in space for this monomer unit in its new trial position, and set the maximal length of this list to $N_{max} = 19$;
(5) we choose an integer random number from the interval [0; 19] and try to form an SB with corresponding neighbours from the list (if this random number is zero then no SB is formed; if this random number is larger than the actual number of real neighbours we reject this trial move; if the chosen neighbour is already involved in some other SB we also reject the trial move).

We accept this trial move according to the usual Metropolis scheme: the acceptance probability is proportional to $\exp(-\mathrm{d}E/T)$, where $\mathrm{d}E$ is the energy difference between the trial and the old states.

The above algorithm satisfies the condition of microscopic reversibility but at the same time it shifts our system to a slightly more compact conformation (see the appendix). To be able to achieve equilibrium and to study properties of the system in a broad interval of temperatures we used the multicanonical ensemble technique [30–33]. The main idea of this

method is to produce a random walk over the states with different energies. This is performed by calculating additional weights for each energy value using the energy histogram obtained during the previous iteration of simulations. Then we run biased MC simulations with those weights and perform reverse reweighting of the simulation data obtained.

The following two remarks are made here concerning the averaging over different primary sequences and the definition of states and transitions between them.

In principle, the averaging of simulation data over the different primary sequences *can* (and in some cases definitely *should*) be performed. However, we did not perform such an averaging systematically here, although we have looked at several particular realizations of randomness. This was done for the following reasons. First, the main aim of the approach we are following is to search for some sequence design procedure which would produce sequences with non-trivial properties (or even probably quite unique ones). We try first to suggest such procedures theoretically and check them in computer simulations, and afterwards think about the possibility of implementing our ideas in the chemical laboratory. Within the framework of this approach one can in principle expect to find that the properties of some particular realization of a sequence are quite different from those averaged over many sequences. This is not expected, of course, for the random (Bernoulli) sequences. Nevertheless, the use of a uniform procedure of comparison of data for just a few particular realizations does have clear justification in our case, because it gives us also an idea about the magnitude of differences between them. This will enable us to locate crucial differences between particular sequences if they appear. Our approach is also supported by some previous experience when important information was obtained from the analysis of differences between the curves for particular primary sequences of designed AB-copolymers [34]. Our second important point is connected with the problem of good averaging over spatial conformations for each particular primary sequence at different temperatures, especially at low ones, taking into account the possibility of formation of different non-trivial spatial intramolecular structures which can depend on the primary sequence. To this end we have used the multicanonical sampling technique which allows good equilibration of chain conformations at different temperatures. Using this technique it is even possible to evaluate the degeneracy of the ground state for a particular sequence [35]. However, this method is time consuming (even for rather short chains consisting of $N = 50$ monomer units which are considered here), which makes it difficult to average over many realizations of randomness.

Our second remark concerns the terminology we are using here and what we understand as 'states', 'transitions' between those states and the 'order' of these transitions. The main points are well known [36, 37]:

(1) there are no phases and phase transitions in a system of finite size;
(2) different spatial structures of a particular polymer chain of finite length can be classified as belonging to some states, i.e. a coil state or globular state, while different intramolecular structures can be found inside dense globules and which can in turn be classified using an appropriate order parameter;
(3) the question of whether the conformational transition between two particular states is the true phase transition can be answered if one extrapolates the behaviour of the system to the thermodynamic limit $N \to \infty$ (the width of the true phase transition goes to zero, $\Delta T \to 0$); we did not perform such an analysis here and emphasize that we speak always about conformational transitions between different intramolecular spatial structures in our particular system of a single chain of finite length (we have to choose an appropriate order parameter to characterize each particular transition);

(4) we consider those transitions to be of 'first-order type' (first-order-like transitions) if they show the formation of the double-peak structure on the histogram of the energy or order parameter, while the shift of a single peak is typical of 'second-order-like' transitions.

The coil–globule transition in polymer chains induced by SBs was studied theoretically some time ago [38, 39]. In the limit of infinitely long chains the phase diagram of temperature versus density of functional units which are able to form SBs, as well as the order of phase transitions between the coil, globule and hairpin, have been investigated. However, we consider in this paper only the case of a rather short AB-copolymer chain, and, moreover, we consider selective interactions only between A- and B-units which can form SBs. Therefore our system cannot be directly compared to the theoretical results either for copolymers with volume interaction or for polymers with SBs. Our main attention concerns the comparison of structural transitions for strongly different primary sequences.

3. Results: coil–globule versus coil–hairpin transition

When the temperature is decreased the chain collapses. To characterize the collapse transition we used the following values: energy E, energy fluctuations, gyration radius R_g, the number of SBs formed N_{SB}. Additionally, we have measured some variables characterizing the formed stems:

(1) the number of individual bonds N_i which do not have any neighbouring SBs;
(2) the number of borders N_b, i.e. the number of SBs which have only one neighbouring SB which belongs to the same stem;
(3) the number of stacked bonds N_s, i.e. the number of SBs having two neighbouring SBs (on the left and on the right) which belong to the same stem (this is the total number of 'internal' SBs involved in the formation of different stems).

The total number of all SBs is the sum of these three values: $N_{SB} = N_i + N_s + N_b$. For three triplets, for example, we have $N_{SB} = 9$, $N_i = 0$, $N_s = 3$, $N_b = 6$.

We studied first the case of the statistically random AB-copolymer (with Bernoulli statistics). In figures 2(a) and (b) we present the temperature dependence of the mean squared gyration radius and of the full energy of a chain. The full chain length is equal to $N = 50$ and we take equal amounts of A- and B-units: $N_A = N_B = 25$. The energy parameter was taken as $\varepsilon_{SB} = -10$ (in kT units) for all different cases considered in this paper. The simulations were performed for three different values of cooperativity parameter $\alpha_{coop} = 0.0$, 0.5 and 1.0. We have presented here the simulation data for only one particular primary sequence; however, we have checked a few other random sequences and found the curves to be similar (see figure 4(b) and discussion below). There are some differences at low temperature which correspond to different ground states but in the transition region the behaviour is quite similar. The system with no cooperativity shows a rather smooth coil–globule transition on decrease of the temperature (circles in figures 2(a) and (b)). The globule is not spherical and its density is much smaller than unity (using only SBs it is impossible to get a dense state). The total number of SBs goes up from 2 to 25 (the highest possible number of SBs in AB-copolymer consisting of 50 monomer units). For cooperativity equal to 0.5 and 1.0 the transition temperature becomes smaller (it drops down from the value $T_{trans} \approx 4.5$ for $\alpha_{coop} = 0$ to $T_{trans} \approx 2.5$ and 2.7 for $\alpha_{coop} = 0.5$ and 1.0 respectively), and the transition itself is sharper. The globular state has larger energy because of positive terms coming from the stem borders. For $\alpha_{coop} = 1.0$ there is even a pretransitional swelling with decreasing temperature which can be understood by taking into account the following argument. At high temperature a few individual bonds

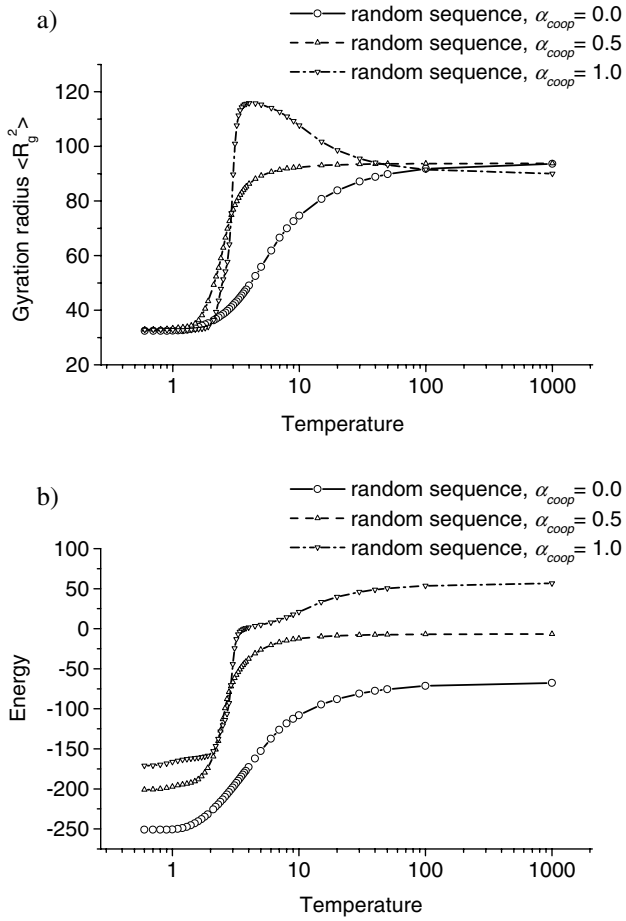


Figure 2. Mean squared gyration radius (a) and mean energy (b) versus temperature for random AB-copolymer for different values of cooperativity parameter α_{coop}.

are always present and cause compactification in comparison to the pure coil state. With decreasing temperatures some of these bonds become unfavourable. The fact that the mean energy is positive for $\alpha_{coop} = 1.0$ is a result of the definitions in table 1. We have also produced histograms of the full energy, and found a simple shift of the single peak in the transition region (no bimodality). At low temperatures there are indications of some conformational transitions (with double-peak histograms) between structures which are quite close to each other in energy (this point requires further investigation).

We have also investigated some specially designed sequences and diblock sequence. The following primary structures were investigated:

(1) random–complementary sequence: this sequence has a statistical 1:1 distribution of A- and B-units in the first half of the chain (Bernoulli statistics) while the second part of the sequence is complementary to the first one (again we have compared the simulation data for several different realizations of random–complementary sequences and found the curves for the temperature dependencies of gyration radius, energy and the number of SBs to be quite close to each other, see figure 4(b) and the discussion below);

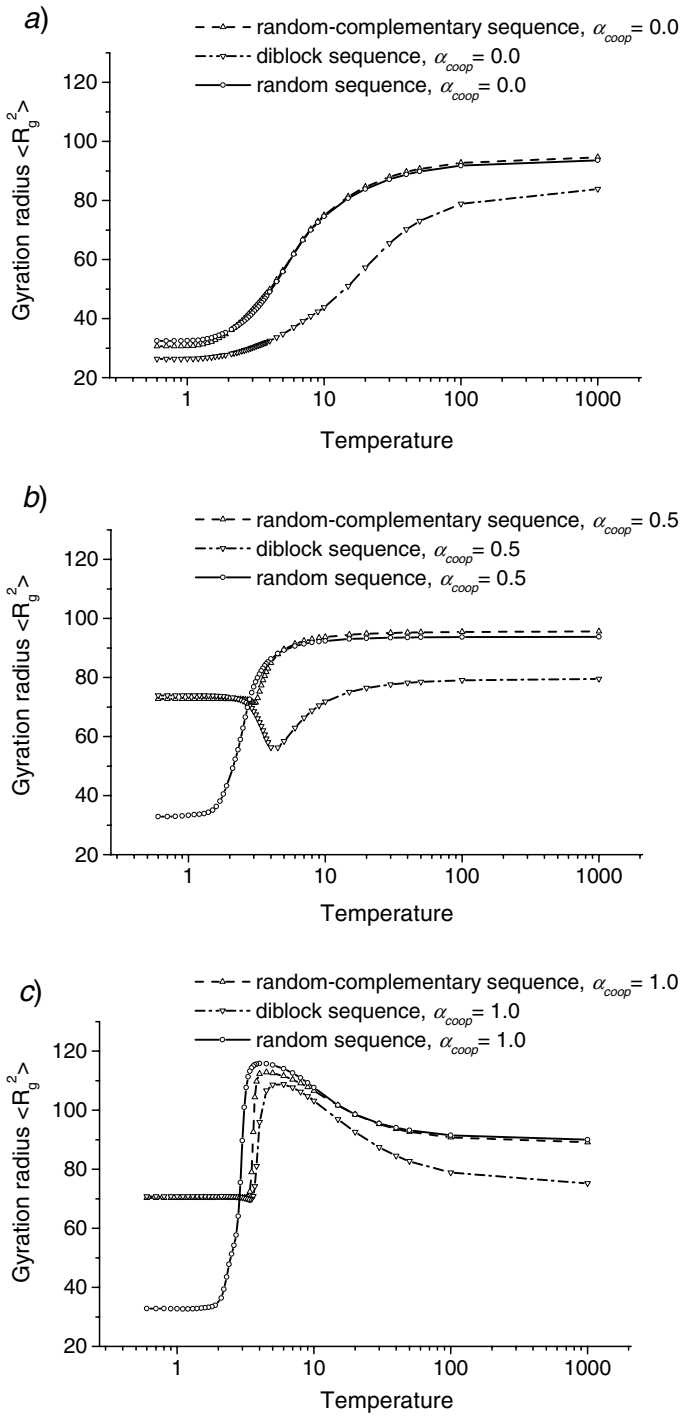


Figure 3. Gyration radius versus temperature for random, diblock and random–complementary sequences for cooperativity parameter $\alpha_{coop} = 0.0$ (a), $\alpha_{coop} = 0.5$ (b), $\alpha_{coop} = 1.0$ (c).

(2) AB-diblock (AB-sequence, which has its first half of the chain consisting of A-units and the other half of B-units).

In figure 3 we present the temperature dependence of the mean squared gyration radius for random, random–complementary and diblock sequences for three values of cooperativity parameter $\alpha_{coop} = 0.0$ (a), 0.5 (b) and 1.0 (c). Figure 4 depicts SB statistics for the secondary structure of the collapsed state of AB-copolymer chains for a cooperativity value equal to 0.5 (figure 4(a)) and 1.0 (figure 4(c)), i.e. the temperature dependence of the number of borders N_b is presented. We have also calculated and analysed the mean energy of the chain and the temperature dependences of the number of individual bonds N_i and the number of stacked bonds N_s for different sequence types (figures are not presented here). In figure 4(b) we show the data for N_b obtained at $\alpha_{coop} = 0.5$ for two particular random sequences and two random–complementary sequences. The number of borders N_b is the variable most susceptible to the particular sequence. These data show that our error bar due to a particular realization of randomness is quite small in the transition region. The ground state (hairpin) is quite unique for the random–complementary sequence while for random copolymers the globules with significant differences in internal structure can be formed at low temperatures.

The analysis of all these data (figures 3 and 4) leads to the following conclusions. Without cooperativity ($\alpha_{coop} = 0$) the properties of random and random–complementary sequences are very close to each other and show remarkable differences from the diblock sequence. The gyration radius for the diblock sequence is smaller than that for random and random–complementary sequences even in the coil state (at high temperatures). The fact that in our model in the coil state the gyration radius of the chain with SB is smaller than that of the chain with only excluded volume interactions is explained in the appendix. At the same time we have found the number of individual SBs to be smaller for the diblock sequence, while the energy is larger (less negative) for this sequence as well (see table 1). This is an apparent contradiction, because the SBs in the diblock sequence can be formed only between monomer units which are more separated from each other along the sequence than in the case of random or random–complementary sequences, and this leads to stronger compactification of the coil state for diblock copolymers.

Upon increasing the cooperativity parameter (let us first discuss the case $\alpha_{coop} = 0.5$) the formation of hairpins is enforced for those chains with primary sequences 'designed' to have a hairpin as the ground state (i.e. diblock and random–complementary sequences). This is not true for random sequences, where the SBs in the compact conformations are mainly the individual ones, i.e. there are many borders, and therefore the energy levels in the low-temperature region are higher. According to the gyration radius data the low-energy conformations for random sequences are closer to the globular ones, but most probably these final conformations are very sequence specific. 'Hairpin-designed' sequences reveal an interesting behaviour, especially chains with the diblock sequence. There are two conformational transitions close to each other, coil–globule and then globule–hairpin: with decreasing temperature the gyration radius decreases and then increases, reaching the value for the hairpin (figure 3(b)). The number of borders (figure 4(a)) initially increases with decreasing temperature, but after achieving a maximum it falls and finally reaches the value 2 on average (hairpin). Such a tendency also takes place for lower values of α_{coop}, even in a more pronounced way (see figure 7(a) below). Random sequences have almost no possibility of decreasing the number of borders, so they persist in the globular state.

The system with a high degree of cooperativity ($\alpha_{coop} = 1.0$) demonstrates a sharp coil–hairpin transition for designed sequences, the effect of an intermediate coil–globule transition being almost absent (figure 3(c)). In all 'designed' sequences after the collapse we have only

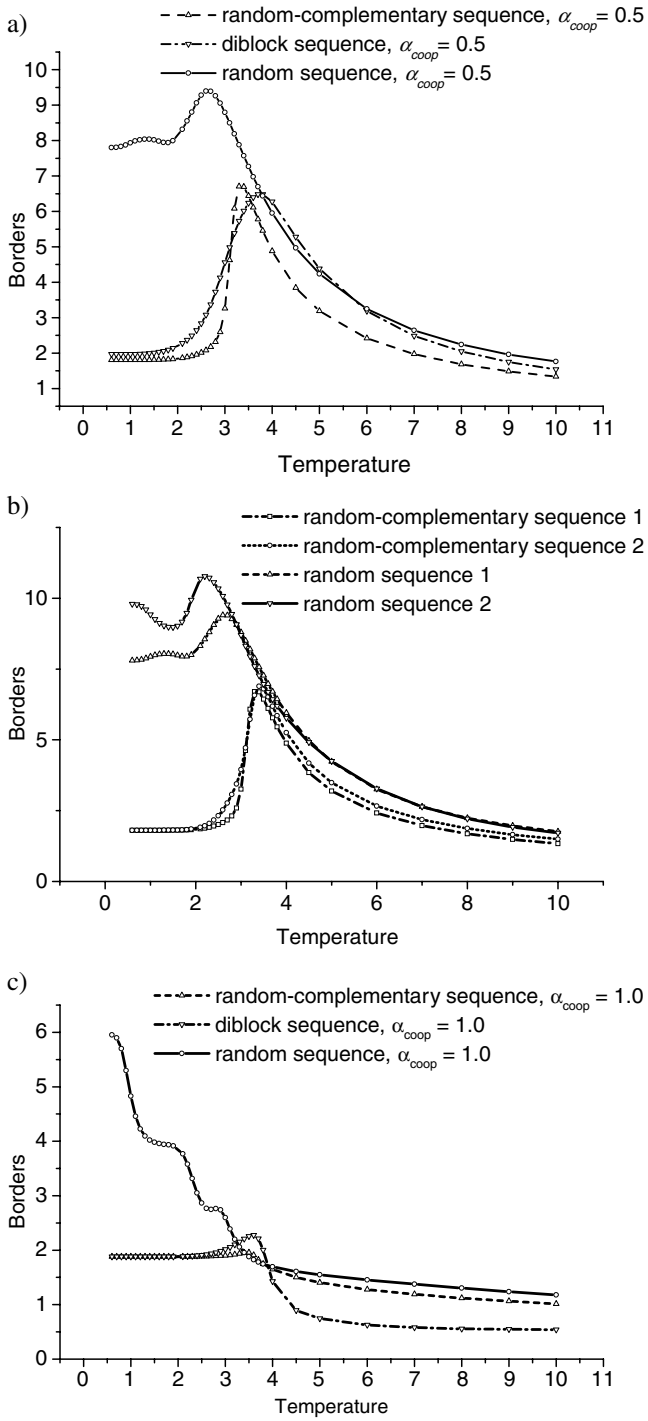


Figure 4. Temperature dependence of the number of borders for different sequences, $\alpha_{coop} = 0.5$ (a) and 1.0 (c). The data for two particular random and two particular random–complementary sequences are shown ((b); $\alpha_{coop} = 0.5$).

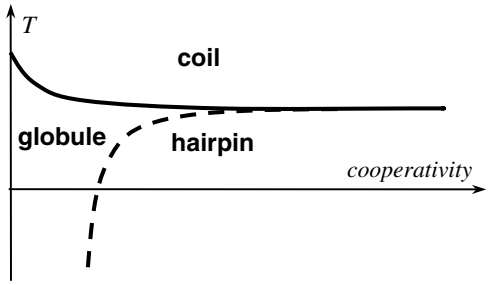


Figure 5. Preliminary diagram of states for variables temperature and cooperativity. The globule–hairpin transition (dashed curve) is present only for diblock and random–complementary sequences but not for the random one. For the random–complementary sequence the globule–hairpin transition line is shifted to lower values of cooperativity (see discussion in the text).

one stem (number of borders equal to 2); in the random sequence case this state is not possible, so we have a kind of globular state with several stems. For some particular sequences one can even distinguish a series of structural transitions between globular states with a different number of stems if this number is considered as an order parameter (see e.g. figure 4(c)). During simulations we had some difficulties with equilibration for this case.

As far as the temperature and sharpness of transitions are concerned, when there is no cooperativity all systems exhibit very smooth coil–globule transitions. There is quite a broad and low peak in energy fluctuations and the transition is of second-order type. When $\alpha_{coop} = 0.5$ we observe decreasing transition temperature and increasing sharpness. At the same time we observe significant differences for different sequences. For the random sequence the decrease of the transition temperature is quite large and the sharpness is the lowest one. The random–complementary sequence reveals a less pronounced decrease of transition temperature and the largest sharpness. The transition for the diblock sequence starts at higher temperature but has lower sharpness, so this sequence forms an ideal hairpin at a temperature which is even lower than that for the random–complementary sequence. Similar behaviour is typical for $\alpha_{coop} = 1.0$, but the transition temperature is slightly higher than in the case of cooperativity 0.5.

In figure 5 we have summarized our results in the form of a schematic diagram of states for variables temperature and cooperativity for a single macromolecule of AB-copolymer with saturating AB-bonds. The transition lines between the coil, globule and hairpin states should be considered more as a guide for the eye because they are based on very few simulation data points and on a general physical sense. The coil, globule and hairpin states are all present for the diblock and random–complementary sequences but the hairpin state is absent for the random sequence. Our data suggest that for the random–complementary sequence the region of the hairpin state is larger, i.e. the globule–hairpin transition line is shifted to the left, in comparison to the diblock sequence. Such a preliminary conclusion is supported by the fact that at $\alpha_{coop} = 0.5$ we observe for the diblock sequence a transition from the coil to hairpin state via the intermediate globular state, but this intermediated globule is not observed for the random–complementary sequence.

We have examined the energy histograms for several temperatures near the transition point. The histograms for $\alpha_{coop} = 0.0$ show a simple shift of a single peak in the course of the coil–globule transition in the AB-copolymer chain of finite length with selective formation of SBs (only between A- and B-units). Therefore, this transition is of second-order type. However, some interesting features can be observed for random–complementary and diblock

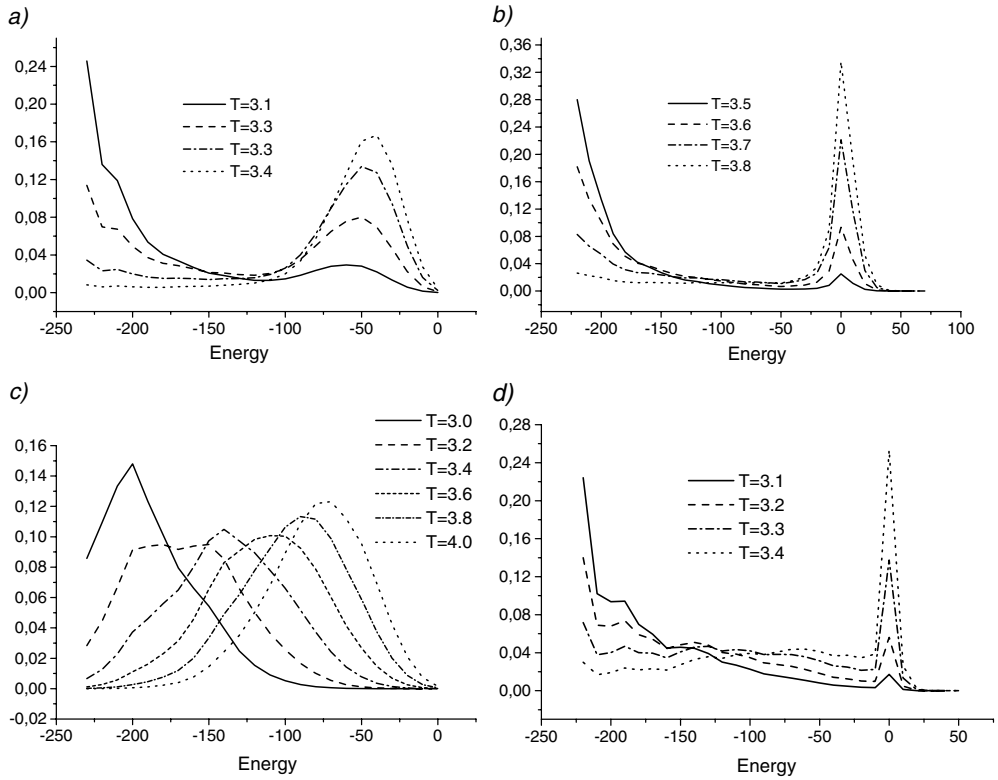


Figure 6. Energy histograms measured at different temperatures for the random–complementary sequence for $\alpha_{coop} = 0.5$ (a) and $\alpha_{coop} = 1.0$ (b) and for the diblock sequence for $\alpha_{coop} = 0.5$ (c) and $\alpha_{coop} = 1.0$ (d).

sequences for $\alpha_{coop} = 0.5$ and 1.0 (figure 6). First, the random–complementary sequence shows a first-order-like transition between the coil and hairpin states for both $\alpha_{coop} = 0.5$ and 1.0 (one can see the bimodality and the decrease of the high-energy peak and a simultaneous increase of the low-energy peak with decreasing temperature). The diblock sequence shows a second-order-like transition between the coil and globule states for $\alpha_{coop} = 0.5$ at temperature $T \approx 3.6$ (broad single peak), although there is quite a clear hint for the first-order-like transition between the globule and hairpin states at lower temperature $T \approx 3.2$. At high cooperativity $\alpha_{coop} = 1.0$ the diblock sequence shows a first-order-like transition between the coil and hairpin states, while there is always some fraction of intermediate middle-energy conformations. For the statistically random sequence (histograms are not shown here) at $\alpha_{coop} = 0.5$ we have a smooth second-order-like coil–globule transition, but then there are indications of a first-order-like transition at even lower temperatures. We explain this phenomenon by assuming some internal readjustments in the globule. An appropriate order parameter can probably be found to describe the transitions between those structures (we have not investigated this problem). Similar indications of several first-order-like transitions were observed for $\alpha_{coop} = 1.0$. We should emphasize here that there are strong finite size effects in our system, and in the thermodynamic limit of large chain length the overall picture of different transitions can change significantly.

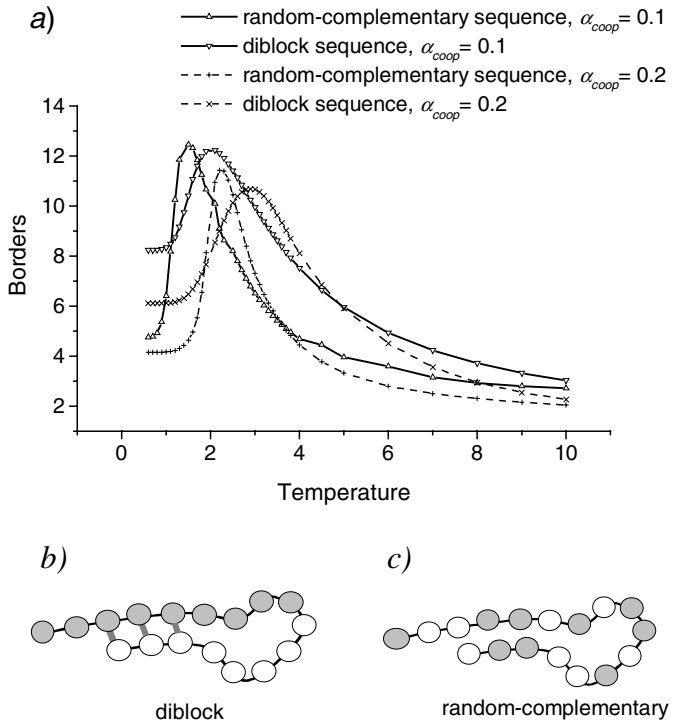


Figure 7. The number of borders for diblock and random–complementary sequences versus temperature at small values of cooperativity parameter (a) and examples of intermediate states for diblock (b) and random–complementary (c) sequences.

What happens with our system at small α_{coop} values? Analysis of the situation near the point where the qualitative behaviour of the system changes (i.e. the coil–hairpin instead of coil–globule transition occurs) may help us to understand differences between diblock and random–complementary sequences. In figure 7(a) our computer simulation data for the temperature dependence of the number of borders for diblock and random–complementary sequences are presented for small values of cooperativity parameter $\alpha_{coop} = 0.1$ and 0.2. One can see that even a rather small value of cooperativity ($\alpha_{coop} = 0.1$) forces the chain with a 'designed' sequence to form hairpin-like structures. The well-pronounced coil–globule and then globule–hairpin transitions are observed. The final structure is not a single hairpin but a set of several rather long stems: four stems on average for the diblock sequence (eight borders) and two stems (four borders) for the random–complementary sequence. The same tendency takes place at higher values of cooperativity ($\alpha_{coop} = 0.2$): the number of stems in the diblock chain (three stems for this particular case) is always higher than in the random–complementary case (two stems). Moreover, for the random–complementary sequence the globular region is narrower and transitions are sharper than in the diblock case. A possible explanation of these differences is the lack of intermediate states in the case of the random–complementary sequence. Indeed, the diblock chain allows some 'impurities' in the secondary structure (because such conformations are not too unfavourable) while the pure hairpin (one stem) for the random–complementary sequence is energetically much more favourable than any imperfections in the structure (see figures 7(b) and (c) where an example of spatial conformation quite close to a perfect hairpin is shown which is completely unfavourable energetically for the

random–complementary sequence). These features can be roughly described by the statement that random–complementary sequences are 'better designed' for the formation of hairpin structure than diblock sequences.

4. Conclusions and outlook

We have presented a new simple model of AB-copolymer chains with the possibility of forming SBs between monomer units of different types, which can be considered as a study of the design of sequences for obtaining simple secondary structures in AB-copolymers.

One important remark should be made here. We consider SBs as a 'first approximation' model for hydrogen bonds, which we keep in mind as our prototype. The properties of real hydrogen bonds are of course much more sophisticated [40]. One of the main features of a hydrogen bond is its dependency on orientation, i.e. such a bond can be formed only if two corresponding functional monomer units are properly oriented relative to each other. This feature is not taken into account in our model. However, our model reflects two other important features of hydrogen bonding in biopolymers, especially in RNA: first that each monomer unit can form only one hydrogen bond; and second that there is an energy gain from the stacking of base pairs in a helix. We start with such a simplified model and claim that it can be used to mimic some important properties of real RNA molecules.

We have found non-trivial features even for this rather simple model. That is, the conformation at low temperatures is changed from the usual globule to a hairpin for some specially designed sequences on increasing the cooperativity parameter. The design of secondary structures consisting of several stems of stacked SBs is our next aim. In general, we can consider a model for sequences with random–complementary sections as a relative successful simple model of RNA-like copolymers.

Acknowledgments

Financial support from the Alexander-von-Humboldt Foundation, Program for Investment in the Future (ZIP), INTAS (project 01-607), and the Russian Foundation for Basic Research is highly appreciated.

Appendix. Proof of the microscopic reversibility condition

The condition of microscopic reversibility can be written in the form

$$p_{12} \cdot \mathrm{e}^{-E_1/kT} = p_{21} \cdot \mathrm{e}^{-E_2/kT} \tag{A.1}$$

where p_{12} is the probability of transition from state 1 (with energy E_1) to state 2 (with energy E_2), and p_{21} is the probability of the backward transition. The transition probability can be represented as

$$p_{12} = \alpha_{12} \cdot \tilde{p}_{12} \tag{A.2}$$

where α_{12} is the probability of choosing state 2 when the system is in state 1 and $\tilde{p}_{12}$ is the probability of accepting the transition from state 1 to state 2. In our model we use the standard Metropolis criteria for $\tilde{p}_{12}$:

$$\tilde{p}_{12} = \mathrm{e}^{-\Delta E/kT} \tag{A.3}$$

where $\Delta E = E_2 - E_1$.

The matrix α_{12} should be symmetric in order to achieve correct distributions for all measured values in a canonical ensemble. Probability α_{12} consists of two parts:

$$\alpha_{12} = \alpha_{12}^{(1)}\alpha_{12}^{(2)} \tag{A.4}$$

where $\alpha^{(1)}$ is the probability of choosing a particular space position for a monomer unit in the course of a trial movement from state 1 to state 2, and $\alpha^{(2)}$ is the probability of choosing a particular neighbour for this unit with which it can form an SB when in state 2. Probability matrix $\alpha^{(1)}$ is symmetric, because we perform trial moves in random directions with a fixed step in space. Therefore, the matrix $\alpha^{(2)}$ should be symmetric as well. It is most natural to choose $\alpha^{(2)} = 1/(N_{max}+1)$, where N_{max} is the maximal number of possible candidates for SB formation and the case of absence of SB is taken into account. For the bond fluctuation model the number of neighbours in a maximally dense state is equal to 25 (without neighbours along the chain), but during simulations of systems with saturating interactions we never observed a value higher than $N_{max} = 19$, so we have chosen $\alpha^{(2)} = 1/20$.

The problem, which follows from this model, is the long lifetime of a particular SB and, as a consequence, the very long relaxation time of the system. At high temperature ($T = 1000$) the mean lifetime of one SB is about 100 MC steps and whole system relaxation time is close to 10^5 MC steps. As can be easily understood from the model, we have effectively a very small coefficient of trial move acceptance, no higher than $1/20$ (in most cases there are almost no suitable neighbours). This leads to some effective attraction between monomer units at high temperatures. The gyration radius in our system is considerably smaller than in the system without SB, in particular at high temperatures. If a monomer unit is involved in the formation of an SB and is surrounded by several neighbours suitable for the formation of another SB, the transition rate to the state without SB will be lower than that to the state with SB (the ratio of these transition rates is proportional to the number of possible SB-neighbours). The existence of several SBs leads to a smaller gyration radius, as in a polymer chain with several stickers. Moreover, this gyration radius will depend on the primary sequence of the copolymer. For the diblock copolymer these stickers are possible only between units which lie far apart from each other along the chain, but for the random sequence the SBs will occur most probably between monomer units which are quite close neighbours along the chain, and this will not lead to a significant decrease of gyration radius.

References

[1] Bryngelson I D and Wolynes P G 1987 *Proc. Natl Acad. Sci. USA* **84** 7524
[2] Yue K and Dill K A 1992 *Proc. Natl Acad. Sci. USA* **89** 4163
[3] Shakhnovich E I and Gutin A M 1993 *Proc. Natl Acad. Sci. USA* **90** 7195
[4] Shakhnovich E I 1994 *Phys. Rev. Lett.* **72** 3907
[5] Klimov D and Thirumalai D 1996 *Phys. Rev. Lett.* **76** 4070
[6] Deutsch J M and Kurosky T 1996 *Phys. Rev. Lett.* **76** 323
[7] Li H, Helling R, Tang C and Wingreen N 1996 *Science* **273** 666
[8] Shakhnovich E I 1997 *Curr. Opin. Struct. Biol.* **7** 29
[9] Dima R I, Banavar J R, Cieplak M and Maritan A 1999 *Proc. Natl Acad. Sci. USA* **96** 4904
[10] Bohr J 2001 *Int. J. Quantum Chem.* **84** 249
Bohr J 2002 *J. Macromol. Sci.* B **41** 787
[11] Pande V S, Grosberg A Yu and Tanaka T 1997 *Biophys. J.* **73** 3192
Pande V S, Grosberg A Yu and Tanaka T 1995 *J. Chem. Phys.* **103** 1
[12] Pande V S, Grosberg A Yu and Tanaka T 2000 *Rev. Mod. Phys.* **72** 259
[13] Khokhlov A R and Khalatur P G 1998 *Physica* A **249** 253
[14] Khokhlov A R and Khalatur P G 1999 *Phys. Rev. Lett.* **82** 3456
[15] Zheligovskaya E A, Khalatur P G and Khokhlov A R 1999 *Phys. Rev.* E **59** 3071

[16] Ivanov V A, Chertovich V A, Lazutin A A, Shusharina N P, Khalatur P G and Khokhlov A R 1999 *Macromol. Symp.* **146** 259
[17] Chertovich A V, Ivanov V A, Lazutin A A and Khokhlov A R 2000 *Macromol. Symp.* **160** 41
[18] Govorun E N, Ivanov V A, Khokhlov A R, Khalatur P G, Borovinsky A L and Grosberg A Yu 2001 *Phys. Rev.* E **64** 040903
[19] Grosberg A Yu and Zhestkov A V 1986 *J. Biomol. Struct. Dyn.* **3** 859
[20] Ubbink J and Odijk T 1995 *Biophys. J.* **68** 54
Odijk T 1996 *J. Chem. Phys.* **105** 1270
[21] Vasilevskaya V V, Khokhlov A R, Kidoaki S and Yoshikawa K 1997 *Biopolymers* **41** 51
[22] Kuznetsov Yu A, Timoshenko E G and Dawson K A 1996 *J. Chem. Phys.* **104** 336
Kuznetsov Yu A, Timoshenko E G and Dawson K A 1996 *J. Chem. Phys.* **104** 105
Kuznetsov Yu A, Timoshenko E G and Dawson K A 1996 *J. Chem. Phys.* **104** 7116
Kuznetsov Yu A and Timoshenko E G 1999 *J. Chem. Phys.* **111** 3744
[23] Nogichi H, Saito S, Kidoaki S and Yoshikawa K 1996 *Chem. Phys. Lett.* **261** 527
Nogichi H and Yoshikawa K 1998 *J. Chem. Phys.* **109** 5070
[24] Ivanov V A, Paul W and Binder K 1998 *J. Chem. Phys.* **109** 5659
Ivanov V A, Stukan M R, Vasilevskaya V V, Paul W and Binder K 2000 *Macromol. Theory Simul.* **9** 488
[25] Higgs P G 2000 *Quart. Rev. Biophys.* **33** 199
[26] Gesteland R F and Atkins J F (ed) 1993 *RNA World* (New York: Cold Spring Harbor Laboratory Press)
[27] Carmesin I and Kremer K 1988 *Macromolecules* **21** 2819
[28] Paul W, Binder K, Heermann D W and Kremer K 1991 *J. Chem. Phys.* **95** 7726
[29] Binder K (ed) 1995 *Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Oxford: Oxford University Press)
[30] Lyubartsev A P, Martsinovskii A A, Shevkunov S V and Vorontsov-Velyaminov P N 1992 *J. Chem. Phys.* **96** 1776
[31] Berg B A and Neuhaus T 1992 *Phys. Rev. Lett.* **68** 9
[32] Escobedo F A and Pablo de J J 1996 *J. Chem. Phys.* **105** 4391
[33] Hansmann U H E and Okamoto Y 1997 *J. Comput. Chem.* **18** 920
[34] van den Oever J M P, Leermakers F A M, Fleer G J, Ivanov V A, Shusharina N P, Khokhlov A R and Khalatur P G 2002 *Phys. Rev.* E **65** 041708
[35] Chertovich A V, Ivanov V A, Khokhlov A R and Bohr 2003 in preparation
[36] Landau L D and Lifshitz E M 1980 *Statistical Physics, Course of Theoretical Physics* vol 5 (Oxford: Pergamon)
[37] Grosberg A Yu and Khokhlov A R 1994 *Statistical Physics of Macromolecules* (New York: American Institute of Physics)
[38] Lifshitz I M, Grosberg A Yu and Khokhlov A R 1976 *Sov. Phys.–JETP* **44** 855
[39] Tanaka F and Ushiki H 1986 *J. Chem. Phys.* **84** 5925
[40] Jeffrey G A 1997 *An Introduction to Hydrogen Bonding* (Oxford: Oxford University Press)