Molecular dynamics simulation of the synthesis of protein-like copolymers via conformation-dependent design

Anatoly V Berezkin\textsuperscript{1}, Pavel G Khalatur\textsuperscript{2,5}, Alexei R Khokhlov\textsuperscript{2,3} and Peter Reineker\textsuperscript{4}

\textsuperscript{1} Department of Physical Chemistry, Tver State University, Tver 170002, Russia
\textsuperscript{2} Department of Polymer Science, University of Ulm, Ulm D-089069, Germany
\textsuperscript{3} Physics Department, Moscow State University, Moscow 119992, Russia
\textsuperscript{4} Department of Theoretical Physics, University of Ulm, Ulm D-89069, Germany

E-mail: khalatur@germany.ru and khokhlov@polly.phys.msu.ru

Abstract. We propose a computer model to describe the synthesis of protein-like copolymers in a polar solvent via irreversible polymerization of hydrophobic and hydrophilic monomers with simultaneous globule formation. In the model, growing chain macroradical, polymerizing monomers, and the preferential absorption of hydrophobic monomers in the core of the globule are taken into account explicitly. The effect of monomer concentrations and reaction rate on the conformational properties and primary copolymer sequences is investigated. We find that, under certain conditions, the resulting copolymers can form protein-like globules with hydrophobic core wrapped in a hydrophilic envelope. Also, a gradient structure of primary sequences of the copolymers is revealed. Such sequences are formed due to both a change in the chain conformation and a continuous redistribution of comonomers between the globule and the solution in the course of the polymerization.
1. Introduction

Heteropolymers, which comprise two or more covalently bonded sequences of chemically distinct monomer species, show a surprisingly rich behaviour both in solution and in bulk [1]–[3]. The simplest examples of two-letter copolymers made up of two different monomer species, denoted by letters A and B, are AB diblock copolymers and regular multiblock copolymer with repeating structure \((AB)_n\). In general, repeated-block copolymers can be written as \((A_iB_jA_kB_l\ldots)_n\), where the indices \(i, j, k, l, \ldots\) denote the number of the corresponding monomer units. More sophisticated distributions of chemically different groups along the chain are characteristic for different random and random-block copolymers, including uncorrelated or ideal random copolymers and the so-called correlated random copolymers. In the former class, the corresponding primary sequences are uncorrelated and correspond to Bernoulli or zeroth-order Markov processes. In the latter class, the correlation in the sequences of both types of A and B segments is defined by means of a first-order Markov process. It is important to emphasize that in both cases the correlations characterizing the distribution of monomer units along copolymer chains decay exponentially. There are, however, such copolymers for which this is not the case.

Recently, the sequence design scheme to generate special primary structures of two-letter quasi-random copolymers with quenched disorder has been proposed in [4, 5]. The idea behind this scheme can be explained as follows. We start with a homopolymer globule stabilized by the attraction between all monomer units. Then a ‘colouring’ procedure is performed. This procedure can be considered as a chemical reaction of a monomer unit with a low-molecular-weight reagent (polymer-analogous transformation). The units in the core of the globule remain ‘red’ and are called hydrophobic (\(H\) units), while the units at the surface of the globule are ‘coloured’ in ‘green’ and are called hydrophilic, or polar (\(P\) units). As a result, monomer units on the surface, instead of being lyophobic, acquire lyophilic properties. Once this primary structure is fixed, attraction between monomer units is removed, and we obtain an \(HP\) copolymer. It is clear that the properties of the resulting copolymer, including its \(HP\) sequence, depend on the bare globular conformation. Such a copolymer was referred to in [4] as a ‘protein-like’ copolymer, because it mimics one of the important features of real globular proteins, namely the possibility of formation of a compact hydrophobic core stabilized by hydrophilic...
Figure 1. Computer-generated movie illustrating the synthesis of a 256-unit protein-like copolymer via polymer-analogous reaction (‘colouring’) from an initially completely hydrophobic homopolymer globule. The polymer chain is depicted by red and green sticks and low-molecular-weight reactant is shown as small yellow spheres. The polymer-analogous reaction between the reactant and the hydrophobic chain units leads to changes in the colour of the chain units. See [4, 5] for more detail.

The statistical properties of protein-like HP sequences were analysed in [6]. It has been shown, both by exact analytical theory and by numerical calculations, that these sequences exhibit long-range correlations of the so-called Levy-flight type [7]; therefore they acquire a certain degree of complexity of the bare globular conformation as a result of a one-step ‘colouring’ procedure [4, 5]. Very recently, it has also been found that the statistics of the arrangement of monomer units along protein-like copolymer chains is distinctly non-Markovian [8] and differ very significantly from the copolymers with random and random-block sequences.

The procedure outlined above was first realized in computer experiments [4, 5]. After the concept of conformation-dependent sequence design was formulated, several possible ways of experimental realization of this concept were proposed. In the literature, the following two main approaches were formulated: (i) chemical modification (‘colouring’) of reactive pendant groups in certain homopolymers and (ii) the formation of a designed protein-like copolymer directly in the course of its synthesis from appropriate comonomers.

The role of ‘colouring’ in real experiments is played by the reaction of a monomer unit with a low-molecular-weight reagent, which converts a hydrophobic unit to a charged or polar group, $\mathcal{H} \rightarrow \mathcal{P}$. In [9], hydrophilization was achieved by grafting of short poly(ethylene oxide) chains to the thermosensitive poly(isopropylacrylamide) backbone. It was shown that grafting to the...
more compact conformation (which occurs mainly from the surface leading to a kind of surface ‘colouring’) is more efficient than random grafting to a coil conformation. Other attempts along the same line were also performed, but in general this method was shown to be rather unreliable, because of the impossibility of stabilizing dense globules in the solution for the time sufficient to implement a polymer-analogous transformation.

The synthesis of protein-like copolymers via copolymerization with simultaneous globule formation was first achieved in [10]. Then similar results were reported in [11, 12] for other pairs of comonomers, thus showing the generality of the results obtained. This approach turns out to be much more robust when compared with the ‘colouring’ procedure. The main idea behind this approach can be explained as follows. Let us assume that we are performing a copolymerization of moderately hydrophobic (thermosensitive) and hydrophilic monomers in aqueous medium. The conditions for copolymerization (e.g. temperature) should be chosen in such a way that when the emerging chain is long enough it forms a globule. Globule formation will lead to a redistribution of monomers: hydrophobic ones should be predominantly concentrated within the globules, whereas hydrophilic monomers will be mainly located in the outside solution. Therefore, when a growing chain end is inside the globular core, mainly hydrophobic units will be added, whereas when the end is located in the outer shell of a globule, the addition of hydrophilic units becomes more preferable. In this way, the protein-like structure with predominantly hydrophobic core and hydrophilic outer envelope should emerge automatically. However, the theoretical foundation of this method is still lacking: the considerations of preferential solvation of monomers presented above are very qualitative, and more exact quantitative investigation is needed. This is just the aim of this paper.

Recently, the experimental approach proposed in [10] has been realized by Berezkin et al [13] in their Monte Carlo simulation study where it was shown that the statistical properties of the sequences obtained as a result of copolymerization in a selective solvent exhibit long-range correlations. The most important advantage of using the computational polymerization procedure introduced in [13] is that it provides detailed information about the composition and conformation of a growing copolymer chain, varying with reaction time, monomer concentrations or their solubility. In this simulation study, however, the polymerizing monomers were taken into account implicitly by the probabilities of attachment of $P$ and $H$ monomers to the growing-chain end. This scheme is a typical Monte Carlo approach.

In the present work, to simulate the irreversible copolymerization of two types of monomers having different solubility and adsorption properties, we will use the method of molecular dynamics. In the model, growing chain macroradical, polymerizing monomers and the preferential absorption of hydrophobic monomers inside the growing globule will be taken into account explicitly.

The rest of the paper is organized as follows. In section 2 we discuss our simulation approach and growth model. We then present our key results and their discussion in section 3. In section 4 we summarize our findings.

2. Model and simulation technique

The growth of polymers from monomers is a complex and fascinating problem. Several factors govern the structure evolution during polymerization. Some of them are reaction kinetics, thermal motion, topological constraints, interactions between polymerizing species, etc. Moreover, the
effect of each factor can change with the evolving polymer conformation. Whether or not two functional units can react is dictated by the thermal energy of the system and reaction chemistry. When functional units collide during random thermal motion, if they have proper orientation and sufficient energy to overcome the activation barrier, they can react to form a single new chemical bond. Two opposite effects control the process of structure buildup during polymerization: diffusion limitations and reaction kinetics. From this point of view, the two limiting cases of polymerization process are well documented, depending on the mobility of monomers.

The polymerization reaction corresponding to the diffusion-controlled limit may be realized, e.g. in a dilute polymer solution with very high viscosity where the chain relaxation time $\tau_{\text{Rel}}$ far exceeds the characteristic time to grow a sufficiently large macromolecule. The simplest computer model, which describes such a polymerization process, is the so-called kinetic growth walk (KGW) model introduced by Stanley and co-workers [14] (see also [15]–[18]).

For the polymerization corresponding to a kinetically controlled regime, the thermal motion of the molecules is implied fast enough so that all reacting species have in principle complete accessibility to each other. For this regime, Flory formulated the principle of equal reactivity of all species participating in the polymerization process [19] (see also [20]–[22]). In reality, however, strong deviations from the principle of equal reactivity are possible [22]. The case of unequal reactivity can be either chemical or physical. Chemical unequal reactivity can be due to the fact that the reactivity of an atomic group depends on the reactive state of other groups belonging to the same molecule. Physical unequal reactivity can arise for a number of different reasons. In particular, when the polymerizing system is composed of two different types of monomers, the interactions between these monomers result in unequal reactivities. Next, spatial constraints may be another reason for unequal reactivity. Finally, diffusion limitations may be a cause for physical unequal reactivity. They can be due to the preferential absorption of one of the types of polymerizing comonomers having different solubility or different affinity to the chemically different sections of a growing copolymer chain. This phenomenon occurs when the conformation of the growing copolymer changes in the course of the polymerization process [22]. This factor will be especially emphasized in the present study.

We simulate the irreversible growth process in the dilute regime under good solvent conditions for hydrophilic (or polar) monomers and under poor solvent conditions for hydrophobic monomers. The polymerization is modelled as a step-by-step chemical reaction of the addition of monomer units to the growing copolymer chain. For the polymerization process, $H$ and $P$ monomers are assumed to be bifunctional and the orientation of individual functional units on the monomers is not distinguished. If a free monomer comes within the prescribed distance $R_{\text{min}}$ (‘reaction radius’) of the end monomer unit of a growing chain, a bond can be formed between the two. Note that, for a radically polymerizing linear polymer, only one of the two ends of the chain is active, and is ready for reacting with free monomers. Also, no polymer rings are allowed to form.

We consider a continuous (off-lattice) model of a copolymer chain that involves two types of monomer units, $H$ (hydrophobic) and $P$ (hydrophilic or polar). To simulate the dynamics of the chain and polymerizing monomers, we employ the method of molecular dynamics (MD). The polymer is treated as a pearl necklace consisting of $N$ segments (monomeric beads), with adjacent beads connected by a rigid rod of fixed length $b$. No additional restrictions are imposed on bond angles (i.e. the angle between two successive bond vectors) other than those naturally arising from the interaction of beads separated by two successive bonds.
Excluded volume interactions between all particles (monomers and chain beads) are modelled via the repulsive part of a Lennard-Jones potential

\[
u(r) = \begin{cases} 
4\varepsilon \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} + \frac{1}{4} \right), & r \leq 2^{1/6}\sigma, \\
0, & r > 2^{1/6}\sigma,
\end{cases}
\]

where \(\varepsilon\) is the energy parameter, \(\sigma\) the diameter of the particle, and \(r\) the distance between particles. Effective hydrophobic attraction is taken into account via the Yukawa-type potential

\[
u_{\alpha\beta}(r) = \begin{cases} 
-\varepsilon_{\alpha\beta} \left( \frac{\sigma}{r} \right) \left[ 1 - \left( \frac{r}{r_c} \right)^2 \right], & r \leq r_c, \\
0, & r > r_c,
\end{cases}
\]

where \(\varepsilon_{\alpha\beta} \in \{\varepsilon_{HH}, \varepsilon_{HP}, \varepsilon_{PP}\}\); \(r_c\) is the cutoff distance for attractive interactions \((r_c = 2.8\sigma)\). In the model, this potential describes the solvent-mediated short-range hydrophobic interactions. The nature of hydrophobic forces is complicated, but for the purpose of the present study it is sufficient to introduce a non-specific short-range attraction using potential (2) in which the (positive) parameter \(\varepsilon_{HH}\) is chosen to promote the preferential absorption of hydrophobic monomers in the core of the globule and the collapse of the copolymer chains. In all the calculations, we set \(\varepsilon_{HH} = 1.61\) (all the energies are measured in units of \(\varepsilon\)). For simplicity, we assume that the solvent molecules surrounding the growing chain are identical to the hydrophilic beads so that \(\varepsilon_{PP} = 0\) and \(\varepsilon_{HP} = \sqrt{\varepsilon_{HH}\varepsilon_{PP}} = 0\). It is clear that for \(\varepsilon_{ab} = 0\), there is no additional interaction between non-bonded particles except that corresponding to the excluded volume potential (1). Potentials (1) and (2) describe the behaviour of both polymerizing monomers and connected monomer units.

Explicitly, no solvent particles are included in the simulations; i.e. the solvent is represented by a dielectric continuum. To simulate solvation effects and the time evolution of the solution in contact with a heat bath of temperature \(T\), we augment the equations of motion by Langevin uncorrelated noise terms [23]. In the MD simulations, the equations of motion of a polymer system in the presence of fixed constraints (the bead-rod model) are

\[
m_i \ddot{r}_i = -\nabla_i U(r) - \sum_{s=1}^{N-1} \lambda_s G_s(r_1, \ldots, r_N) - \Gamma_i \dot{r}_i + \eta_i, \quad i = 1, 2, \ldots, N + n,
\]

with the constraints of fixed bond lengths

\[
G_s(r_1, \ldots, r_N) = \frac{1}{2} [(r_{s+1} - r_s)^2 - b^2] = 0, \quad s = 1, 2, \ldots, N - 1,
\]

where \(N\) is the number of beads in the chain \((N = N_H + N_P)\), \(n\) denotes the number of non-bonded monomers in the system \((n = n_H + n_P)\), \(m_i = 1\) is the mass of the chain bead or monomer \(i\) with position vector \(r_i\), \(-\nabla_i U(r)\) represents the total non-bonded forces on the particle \(i\) (\(U\) is the total potential energy), \(b = \sigma\) is the bond length, \(\lambda_s\)'s are the Lagrange multipliers, \(\eta_i\) describes the random force of the heat bath acting on particle \(i\), and \(\Gamma\) takes into account
the viscosity of the solvent. The summation in equation (3) is performed over all the $(N - 1)$ bonds of an $N$-unit polymer chain. This term represents the forces, $\nabla_i \sum_{j=1}^{N-1} \lambda_j G_i(r_i, \ldots, r_N)$, which are due to the bond reactions. Each of the $i$th geometrical constraint (bond) is associated with exactly one unknown $\lambda_j$. In equation (3), $\Gamma_0$ and $\Gamma_i$ are connected through the fluctuation–dissipation theorem, $(R_{ai}(0) \cdot R_{ai}(t)) = 2 \Gamma_i k_B T \delta(t)$, $(a = x, y, z; T$ is the reference temperature) and ensures that the temperature is kept constant. The equations of motion (3) in conjunction with equation (4) are solved iteratively using a Newtonian iteration procedure with the time step $\Delta t = 0.01 \sigma \sqrt{m/\epsilon}$ (see [23]–[25] for more technical details). The unit of time in the simulation is defined by $\tau = \sigma \sqrt{m/\epsilon}$. In the following, the reference temperature $T$ (measured in units of $\epsilon/k_B$) is fixed at $T = \epsilon/k_B$.

We take the parameter $\Gamma$ to be dependent on solvent-accessible surface areas (SASA). To find the values of SASA for a given configuration, we perform an analytical computation of the surface areas $A_i$ for each specified particle [26]. Having $A_i$, one can define $\Gamma_i$ as $\Gamma_i = \Gamma_0 A_i/A_{max}$, where $A_{max}$ is the maximum solvent-accessible surface area of a particle for the model under study and the reference value of $\Gamma_0$ is taken to be equal to $0.5 \sigma^{-1}$. The weighting factor $A_i/A_{max}$ represents the degree of exposure of the particle $i$ to the solvent. When the value of SASA for a given particle is zero, the frictional and random forces are zero and the Langevin equation (3) reduces to Newton’s equation of motion. For the chain in a globular conformation, this typically happens when the monomer unit is located in the core of a globule. In contrast, a monomer unit located at the globular surface is strongly solvated; it means that $A_i$ should be close to $A_{max}$ and, as a result, the value of $\Gamma_i$ is close to its reference value $\Gamma_0$.

Our algorithm simulates a copolymerization reaction in a three-dimensional (3D) cube utilizing periodic boundary conditions. Initially, $n$ freely diffusing monomers (hydrophobic and hydrophilic) are placed randomly in the cube. The total number density of the monomers is fixed at $\rho = 0.01$. After a monomer is attached to the growing chain, a new free monomer of the same type is placed randomly in the box, checked for overlaps with already placed particles. Then this monomer gets a velocity of a given absolute value with a randomly chosen direction. The concentration of monomers in solution does not depend on their conversion and, therefore, always remains a constant. In most of the runs discussed below, the model system involved 20 and 80% $H$ and $P$ monomers, respectively. The reaction radius was set to $R_{min} = b \pm \delta$, where $\delta = 0.01 \sigma$, i.e. a monomer was allowed to be attached to the end bead of the growing chain when the corresponding distance $r$ was in the range $b - \delta \leq r \leq b + \delta$. Note that for the chain lengths simulated in this study ($N \leq 512$), we in fact dealt with a dilute polymer solution.

The usual radical polymerization process consists of the following three phases: initiation, chain propagation and termination. Since we put only one radical initiator into the reaction mixture, spontaneous chain-terminating reactions, such as radical combination and disproportionation, were excluded. This situation takes place when the concentration of radical species in a polymerization reaction is very small compared with other reactants. The polymerization is modelled as a sequence of the alternating steps: (i) the addition of a new monomer unit to the growing chain and (ii) the relaxation of the chain with the current length $N_t$. Polymerization occurs without crossing energy barriers and is determined only by the local concentration of reactive monomers near the ‘active’ end of the macroradical. In the course of the chain relaxation, the average number of $H$ and $P$ monomers, $n_H(\tau)$ and $n_P(\tau)$, around the chain end in the reaction volume $\frac{4}{3} \pi R_{min}^3$ are calculated. Since the probability of a reaction is proportional to its rate, the probability for $H$ and $P$ monomer binding can be
written as

\[
V_H = \frac{n_H(\tau)}{n_H(\tau) + n_P(\tau)}, \quad V_P = \frac{n_P(\tau)}{n_H(\tau) + n_P(\tau)}, \quad V_H + V_P = 1. \tag{5}
\]

Thus, during the simulation, we find the current position of the ‘active’ end of the chain \([x_N(\tau), y_N(\tau), z_N(\tau)]\) and add to it a new monomer unit, according to the relative probability for the occurrence of the monomer \(H\) or monomer \(P\). It is clear that, in our polymerization model, the values of \(V_P\) and \(V_H\) should depend on the position of the growing chain end: if it is located in the region occupied mainly by hydrophobic species, the value of \(V_H\) dominates; otherwise, \(V_P\) dominates. A random number uniformly distributed between 0 and 1 is generated and then compared with the probability \(V_P\); if this random number is less than the value of \(V_P\), the \(P\) monomer unit is added to the chain end; otherwise the \(H\) monomer unit is added. During this and subsequent stages of the simulation, the new terminal bead interacts with intramolecular potentials (1) and (2), depending on the kind of the bead, \(H\) or \(P\). After that, the relaxation of the growing \(N\)-unit macroradical begins; this takes \(\tau_R\) MD steps. The time \(\tau_R\) that is allowed for this ‘annealing’ between attachments is used as a control parameter to produce characteristically different chain conformations as discussed below. Note that since in our model the concentration of the monomers is fixed, the quantity \(\tau_R\) does not depend on the concentration.

The value of \(\tau_R\) can be considered as a typical reaction time that, for simplicity, is considered as a constant. In reality, however, this time is a stochastic variable determined by random number \(\eta\) distributed uniformly over the unit interval:

\[
\tau_R \propto \ln\left(\frac{1}{\eta}\right) \frac{1}{w}, \tag{6}
\]

where \(w\) is the reaction rate. Generally speaking, the assumption of a constant reaction time does not affect the polymer properties as long as \(\tau_R\) is greater than typical times of other processes taking place in the system [13]. There are two such processes: monomer diffusion with a typical time scale \(\tau_D\) and chain relaxation with a typical time \(\tau_{Rel}\). For a kinetically controlled polymerization, the condition \(\tau_R > \tau_D\) is sufficient. However, the conformation-dependent design imposes an additional condition: \(\tau_R > \tau_{Rel}\) [8, 13]. In this case, the reaction time should exceed not only the diffusion time \((\tau_D)\) but also the relaxation time \((\tau_{Rel})\) of the \(N\)-unit chain, which by itself could be much higher than \(\tau_D\) for sufficiently long chains. In other words, we should set [8, 13]

\[
\tau_D < \tau_{Rel} < \tau_R. \tag{7}
\]

The value of \(\tau_R\) was varied from 100 to 20 000 integration steps. During the reaction, information was tabulated at various checkpoints for later analysis. Copolymerization was ended when the chain length reached \(N = 512\) units. After the termination of the growing chains, their average conformational properties were calculated during \(10^9\) MD time steps. Also, the primary structures of the synthesized copolymer chains were analysed. The distributions of monomer units are also calculated as a function of the distance from the centre of mass. To gain better statistics, approximately \(10^3\) copolymer chains for each set of the parameters were generated and the required average characteristics were found. The typical CPU time needed for generating one copolymer chain of 512 monomer units on a 1.8 GHz Pentium IV processor is approximately
3. Results and discussion

3.1. General trends in chain growth

It should be emphasized that, for all samples that are considered in the present study, the resulting copolymer chains tended to be compact due to strong attraction between the \( \mathcal{H} \) monomer units. To give a visual impression of the simulated system, figure 2 presents a typical snapshot picture of an 512-unit copolymer chain synthesized at \( \tau_R = 10^4 \). Figure 3 presents the corresponding movie showing the evolution of the growing chain. At this stage we are interested in qualitative effects only.

Already from these data one can easily see that, using the polymerization mechanism proposed in this work, we can indeed end up with a protein-like copolymer having a dense hydrophobic core surrounded by a hydrophilic shell. In general, the same core–shell structure is observed for protein-like globules obtained through the ‘colouring’ procedure (see figure 1 and [4, 5]).

3.2. Reaction time (\( \tau_R \))

Strictly speaking, our kinetic model gives correct results only in the limit, when \( \tau_R \gg \tau_D \) and \( \tau_R \gg \tau_{Rel} \) [8, 13]. By varying \( \tau_R \), one can find such conditions, when this limiting regime is
established. To check this point, we carried out a series of simulations for different values of $\tau_R$. In particular, we calculated the average fractions of $\mathcal{H}$ monomer units in the resulting copolymer, $\phi_H$, its mean-square radius of gyration, $R^2_g$, and the partial mean-square radii of gyration, $R^2_{g_H}$ and $R^2_{g_P}$, found separately for $\mathcal{H}$ and $\mathcal{P}$ segments. All these quantities are shown in figures 4 and 5 as a function of $\tau_R$.

As seen from figures 4 and 5, at small $\tau_R$, the average composition and the partial mean-square radii of gyration correspond to a random copolymer for which, obviously, one has: $\phi_H N_H / (N_H + N_P) = n_H / (n_H + n_P) = 0.2$, $R^2_{g_H} \approx R^2_{g_P} \approx R^2_{g}$. In this case, the probabilities of attachment of $\mathcal{H}$ and $\mathcal{P}$ monomers are determined only by the concentrations of the monomers in the reaction mixture and the polymerization process is governed by the principle of equal reactivity of Flory [19]. On the other hand, as the $\tau_R$ value is increased, these quantities deviate from those expected for a random copolymerization and asymptotically approach constant values. Such a behaviour corresponds to the conformation-dependent design regime for which inequality (7) becomes true [13]. Therefore, for sufficiently large reaction times, we actually deal with the conformation-dependent copolymer design.

As seen from figure 4, the fraction of $\mathcal{H}$ monomer units in the resulting copolymer increases with $\tau_R$. This behaviour has a simple explanation. At small $\tau_R$, the copolymer chain grows very fast so that chain conformation cannot relax completely and, as a result, $\mathcal{H}$ units cannot form the hydrophobic core which absorbs $\mathcal{H}$ monomers from the solution. In this case, therefore, only usual random copolymerization is possible. On the other hand, for sufficiently large $\tau_R$, owing to the preferential absorption of $\mathcal{H}$ monomers from the solution by the hydrophobic core of the growing globule, the possibility of adding $\mathcal{H}$ monomer units increases and their fraction in the

Figure 3. Computer-generated movie illustrating the evolution of the growing chain during copolymerization of hydrophobic and hydrophilic monomers with simultaneous globule formation. The $\mathcal{H}$ and $\mathcal{P}$ units are shown as red and yellow spheres, respectively. The growing chain end is depicted as a larger green sphere. For visual clarity, the polymerizing monomers are not shown.
Figure 4. Average fractions of $H$ monomer units in the resulting copolymer versus $\tau_R$. 

Figure 5. Mean-square radius of gyration, $R_g^2$, and partial mean-square radii of gyration, $R_{gH}^2$ and $R_{gP}^2$, found separately for $H$ and $P$ segments as a function of reaction time. 

copolymer also increases. Visible differences between the partial mean-square radii of gyration for $H$ and $P$ units (figure 5) reveal the fact that distinct segregation of the two types of monomer units in the core–shell conformation takes place. In particular, the value of $R_{gH}^2$ characterizing the relative size of hydrophobic core becomes smaller than the analogous value $R_{gP}^2$ corresponding...
to the hydrophilic envelope. Such a behaviour corresponds to the conformation-dependent design regime for which we observe a strong deviation from the principle of equal reactivity of Flory. Indeed, although the concentration of $H$ monomers in solution is lower when compared with $P$ monomers and the reactivities of these monomers are equal, the copolymer synthesized in this regime is preferentially enriched by hydrophobic monomer units. The deviation from Flory’s principle is due to the interplay between monomer mobilities and hydrophobic attraction, on the one hand, and steric hindrances caused by the polymer chain in a globular conformation, on the other [22]. If monomers react much faster than their diffusion around the reaction chamber, they almost always react in small neighbourhoods, and usually this leads to random connections, i.e. to the formation of a trivial statistical copolymer. However, if monomers react very slowly, the growing copolymer chain has time to attach a sufficiently large number of hydrophobic monomers and to form a compact conformation with a dense hydrophobic core; as a result, an interface arises between two regions in which either hydrophobic species or hydrophilic species dominate. Spatial constraints (or steric hindrances) at the interface and inside the hydrophobic region limit the diffusion, thus preventing the intermix of the species. Due to this, when the growing chain radical is located inside the hydrophobic globular core, it attaches mainly $H$ monomers strongly absorbed here; being outside the core, it gathers mainly $P$ monomers. In a sense this process is similar to a random walk near an infinite plane when the walk crossing the plane changes its ‘colour’. It is well known that the ‘colour distribution’ in this walk demonstrates a non-trivial (Levy-flight) statistics [6, 7]. We would also like to emphasize that, in our model, the type of the unit attached to the growth centre during the simulation, under kinetic control, is determined by the conformation and primary structure of the growing chain as a whole, and not by the local concentration of reactive monomers near the ‘active’ end of macroradical. Such a co-operativity is an additional cause for the physical unequal reactivity of monomers.

It is necessary to stress that, in principle, the synthesis of random and protein-like copolymers proceeds simultaneously. However, the relative contribution of each of these processes significantly depends on the balance between the velocity of the chain relaxation (which is closely related to copolymer composition) and a typical reaction time. This statement is illustrated by the results shown in figure 6, where we present the fraction of different chains, $W(\phi_P)$, as a function of chemical composition, $\phi_P = N_P / (N_H + N_P)$, for a few values of $\tau_R$. It is seen that at small and large times $\tau_R$, there is only one type of copolymer, which can be attributed to random and protein-like copolymers, respectively. In other words, the $W(\phi_P)$ function is unimodal. On the other hand, at intermediate times ($\tau_R = 5 \times 10^3$ time steps), there are two types of copolymers, one of which is preferentially a hydrophobic protein-like copolymer and the other is preferentially a polar random copolymer.

From the simulations we can conclude that there is a strong correlation between the composition of the growing copolymer, which determines the chain conformation, and the type of the monomer, which is attached. It is clear that the chain with high content of $H$ segments can form compact globular structures with absorbed $H$ monomers more easily, and this further facilitates the reactions with $H$ monomers. In contrast, the conformation of the chain with high content of polar units is more expanded and in this case the reaction with $P$ monomers is more probable because their solution concentration is higher.

Apparently, the critical reaction time, when the chain relaxation competes with the copolymerization, is $\tau_R = 5 \times 10^3$. At $\tau_R > 5 \times 10^3$, conformation-dependent design proceeds preferably when compared with random copolymerization.
3.3. Copolymer structure at large $\tau_R$

For synthesized copolymers, we have studied the variation of the average $\mathcal{H}\mathcal{P}$ composition at different content of $\mathcal{H}$ and $\mathcal{P}$ monomers in the reaction mixture. Figure 7 shows the fraction of connected $\mathcal{P}$ units, $\phi_P$, as a function of the composition of the initial monomer mixture $\phi_P^0 = n_P/(n_H + n_P)$ at $\tau_R = 10^4$. It is clear that the equality $\phi_P = \phi_P^0$ corresponds to a random (statistical) copolymer. As seen from figure 7, the chemical composition of the resulting copolymer strongly deviates from the statistical composition. Also, we can conclude that a 1:1 $\mathcal{H}\mathcal{P}$ composition is obtained for $\phi_P^0 \approx 0.8$; i.e. the synthesis of such copolymers is possible only at large excess of polar monomers in the reaction mixture and the acceptable 1:1 $\mathcal{H}\mathcal{P}$ comonomer composition is very narrow. Thus the change in the solution concentration of monomers allows a variation of the composition of synthesized copolymers over a wide range. Below, we will discuss only the results obtained at $\phi_P^0 = 0.8$.

In figure 8 we present the radial distribution of $\mathcal{H}$ and $\mathcal{P}$ monomer units in the globules synthesized at $\tau_R = 10^4$, where the centre of mass of the globular core is taken as the origin. The radial distribution of free $\mathcal{H}$ and $\mathcal{P}$ monomers around the centre of the globule at the same $\tau_R$ value is shown in figure 9.

The inspection of figure 8 demonstrates that the copolymer chains form a core–shell structure in which the $\mathcal{H}$ and $\mathcal{P}$ segments are well segregated. An analogous conclusion has been drawn from the data presented in figures 2 and 3. From figure 9 it is seen that the distribution of free $\mathcal{H}$ and $\mathcal{P}$ monomers inside the globule is also inhomogeneous. In accordance with our suppositions, an excess of hydrophobic species in the globular core (at $r \lesssim 5\sigma$) arises (figures 8 and 9), while in the solution (at $r > 6\sigma$) an excess of $\mathcal{P}$ monomers is observed (figure 9). Hence, at the centre of the globule, the active chain radical reacts with $\mathcal{H}$ monomers, while $\mathcal{P}$ monomers are joined preferentially near the surface of the globule.

---

**Figure 6.** Fraction of copolymer chains with different $\mathcal{H}\mathcal{P}$ composition, $W(\phi_P)$, as a function of chemical composition, $\phi_P = N_P/(N_H + N_P)$, for a few values of $\tau_R$. 

**Figure 7.** Radial distribution of copolymer chains near the centre of the globule synthesized at $\tau_R = 10^4$. The centre of mass of the globular core is taken as the origin. The radial distribution of free $\mathcal{H}$ and $\mathcal{P}$ monomers around the centre of the globule at the same $\tau_R$ value is shown in figure 9.
Figure 7. Fraction of connected $P$ units as a function of the composition of the initial monomer mixture $\phi^0_P = n_P / (n_H + n_P)$ at $\tau_R = 10^4$.

Figure 8. Radial distribution of $H$ and $P$ chain beads in the globules synthesized at $\tau_R = 10^4$. The centre of mass of the globular core is taken as the origin.

It is necessary to note that at the core–shell boundary (at $r \approx 4.5 \sigma$) $H$ monomers form an adsorption layer. The same feature has been observed in [13], for the monomer distribution where we have carried out the simulation of a homopolymer globule enclosed in a periodic box with hydrophobic monomers.
Figure 9. Radial distribution of free $H$ and $P$ monomers in the globules synthesized at $\tau_R = 10^4$. The centre of mass of the globular core is taken as the origin.

3.4. Analysis of copolymer sequences

Let us now consider some statistical properties of the primary sequences of the obtained copolymers. To this end, we calculated the average numbers of $H$ and $P$ blocks and the average lengths of these blocks, $L_H$ and $L_P$. By the average block length we mean here the average number of $H$ (or $P$) monomer units in a chemically homogeneous chain segment built up from these units. The values of $L_H$ and $L_P$ are shown in figure 10 as a function of the reaction time $\tau_R$.

As seen, the length of hydrophobic blocks increases with $\tau_R$ while the value of $L_P$ is a decreasing function of $\tau_R$. To explain this effect, it should be kept in mind that if $\tau_R$ is very small, when conformational equilibrium is lost, chain growth resembles a random walk connecting free monomers occurring in the capture radius of the growth centre. In other words, in this case the growing chain segment does not have enough time to get a compact conformation. This enhances the probability of chemical reactions in the polar regions of the forming copolymer and, at the same time, decreases the probability of penetration of the chain end-radical into hydrophobic regions. As a result, hydrophilic monomers occur in the reaction volume more frequently.

For comparison, we show in the same figure the data obtained for a purely random 1:1 sequence. The average block lengths for the synthesized copolymers are found to be greater than those for a random copolymer with 1:1 $H/P$ composition for which $L_H = L_P = 2$. This fact can be explained by ‘tuning’ the sequence of the growing chain to globule size. The changes in average block lengths are explained mainly by the change in the copolymer composition. Note also that the average block length distributions for the synthesized copolymers are significantly broader than those for a random copolymer.

We found that, owing to the specific mechanism of the chain propagation, the synthesized copolymers have a gradient primary structure, i.e. the $H/P$ composition changes along the
Figure 10. Average lengths of hydrophobic and hydrophilic blocks, $L_H$ and $L_P$, as a function of the reaction time $\tau_R$.

growing chain during copolymerization. This result follows from the analysis of the distribution of monomer units along the chain. In figure 11, we present the average fraction of $H$ units in the chain, $\phi_H(i)$, as a function of monomer unit number, $i$, for a few values of $\tau_R$. It is seen that the function $\phi_H(i)$ is practically constant at small $\tau_R$ when a random copolymer is synthesized. On the other hand, when the value of $\tau_R$ becomes sufficiently large, the local $H$ composition changes along the growing chain during copolymerization. Short chains do not contain a sufficiently large number of hydrophobic units to form a core–shell conformation and the growth of these chains is very similar to that observed for random copolymerization. The composition of these chains is also close to the composition of the initial monomer mixture. Then, as the chain becomes longer and the number of connected $H$ units increases, a dense hydrophobic core can be formed. This core can intensively absorb $H$ monomers and their fraction in the resulting copolymer increases. In this case, the probabilities of monomer addition expressed by equation (5) are also not constant due to the changes in the ratio between the volumes of the hydrophobic core and the polar shell during the chain growth. The gradient structure becomes more pronounced on increasing the reaction time.

4. Conclusion

In this work, we have carried out molecular dynamics simulations to describe the process of irreversible solution copolymerization with simultaneous globule formation. The presence of polymerizing monomers in the reaction mixture and the preferential sorption of hydrophobic monomers in the core of globule was explicitly taken into account. Our approach is similar to that first experimentally realized by Lozinsky et al [10] (see also [11, 12]). Although in this study we have only dealt with the simplest model of copolymerization in a selective
Figure 11. Average fraction of $H$ units in the chain as a function of bead number for a few values of $\tau_R$.

solvent, nevertheless, we believe that the results obtained from this simulation will be helpful for understanding the nature of the experimentally observed features and for providing ideas for further experiments. In particular, we have monitored the formation of a two-letter quasi-random copolymer with quenched primary structure, which demonstrates several non-trivial features. The effect of monomer concentrations and reaction rate on the conformational properties and primary copolymer sequences has been investigated in detail. We have found that, under certain conditions, the resulting copolymers can form protein-like globules with hydrophobic core wrapped in a hydrophilic envelope. An analogous conclusion has been drawn in our previous theoretical study [13] where the copolymerization in a selective solvent was simulated without explicit monomers. Also, we have found that the synthesized copolymers have a gradient structure of primary sequences. From our simulation it follows that such sequences are formed due to both a change in the chain conformation and a continuous redistribution of comonomers between the globule and the solution in the course of the polymerization. Finally, for the system under study we have observed a strong deviation from the principle of equal reactivity of Flory [19]. Although the concentration of hydrophobic monomers in the simulated system was lower when compared with hydrophilic monomers and the reactivities of both monomers were equal, the irreversible copolymerization with simultaneous globule formation leads to a copolymer that is preferentially enriched by hydrophobic monomer units. The basis for this effect is the interplay between monomer mobilities and hydrophobic attraction, on the one hand, and steric hindrances caused by polymer chain in a globular conformation, on the other.

Acknowledgments

We thank S I Kuchanov for stimulating discussions. The financial support from the Alexander-von-Humboldt Foundation, Program for Investment in the Future (ZIP), INTAS (project #01-607) and Russian Foundation for Basic Research is highly appreciated. AVB acknowledges the ‘Deutscher Akademischer Austauschdienst’ (DAAD) for financial support.
References

[1] Fredrickson G H and Bates F S 1996 Ann. Rev. Mater. Sci. 26 501
[2] Bates F S and Fredrickson G H 1996 Ann. Rev. Phys. Chem. 41 525
[3] Davis K A and Matyjaszewski K 2002 Statistical, Gradient, Block and Garnet Copolymers by Controlled/Living Radical Polymerizations (Berlin: Springer)
[4] Khokhlov A R and Khalatur P G 1998 Physica A 249 253
[5] Khalatur P G, Ivanov V I, Shusharina N P and Khokhlov A R 1998 Russ. Chem. Bull. 47 855
Khokhlov A R and Khalatur P G 1999 Phys. Rev. Lett. 82 3456
[6] 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(R)
[7] Shlesinger M F, Zaslavskii G M and Frisch U (ed) 1996 Levy Flights and Related Topics in Physics (Lecture Notes in Physics vol 450) (Berlin: Springer)
[8] Kuchanov S I and Khokhlov A R 2003 J. Chem. Phys. 118 4672
[9] Virtanen J, Baron C and Tenhu H 2000 Macromolecules 33 336
Virtanen J and Tenhu H 2000 Macromolecules 33 5970
Virtanen J, Lemmetiyinen H and Tenhu H 2001 Polymer 42 9487
[10] Lozinsky V I, Simenel I A, Kurskaya E A, Kulakova V K, Grinberg V Ya, Dubovik A S, Galaev I Yu, Mattiasson B and Khokhlov A R 2000 Rep. Russ. Acad. Sci. 375 637
Lozinsky V I, Simenel I A, Kurskaya E A, Kulakova V K, Grinberg V Ya, Dubovik A S, Galaev I Yu, Mattiasson B and Khokhlov A R 2000 Dokl. Chem. 375 273
[11] Wahlund P-O, Galaev I Yu, Kazakov S A, Lozinsky V I and Mattiasson B 2002 Macromol. Biosci. 2 33
Lozinsky V I et al 2003 Macromolecules 36 7308
[12] Siu M-H, Zhang G and Wu C 2002 Macromolecules 35 2723
[13] Berezkin A V, Khalatur P G and Khokhlov A R 2003 J. Chem. Phys. 118 8049
[14] Majid I, Jan N, Coniglio A and Stanley H E 1984 Phys. Rev. Lett. 52 1257
[15] Herrmann H J 1984 Kinetics of Aggregation and Gelation ed F Family and D P Landau (Amsterdam: North-Holland)
[16] Lyklema J W 1986 Fractals in Physics ed L Pietronero and E Tosati (Amsterdam: North-Holland)
[17] Lucena L S, Araújo J M, Tavares D M, da Silva L R and Tsallis C 1994 Phys. Rev. Lett. 72 230
[18] Neves U P C and Onody R N 1995 Physica A 218 1
[19] Flory P J 1953 Principles of Polymer Chemistry (Ithaca, NY: Cornell University Press)
[20] Gordon M 1962 Proc. R. Soc. A 268 240
[21] Miller D R and Macosko C W 1976 Macromolecules 9 199
Ziff R M and Stell G 1980 J. Chem. Phys. 73 3492
[22] Kuchanov S I, Keshtov M L, Khalatur P G, Vasnev V A, Vinogradova S V and Korshak V V 1983 Makromol. Chem. 184 105
[23] Allen M P and Tildesley D J 1990 Computer Simulation of Liquids (Oxford: Clarendon)
[24] Andersen H C 1983 J. Comput. Phys. 52 24
[25] Khalatur P G, Balabaev N K and Pavlov A S 1986 Mol. Phys. 59 753
[26] Wesson L and Eisenberg D 1992 Prot. Sci. 1 227
Augspurger J D and Scheraga H A 1996 J. Comput. Chem. 17 1549