Design of random copolymers with statistically controlled monomer sequence distributions via Monte Carlo simulations

James J. Semler\textsuperscript{a} and Jan Genzer\textsuperscript{b}

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27695-7905

(Received 23 January 2006; accepted 11 May 2006; published online 5 July 2006)

We use Monte Carlo simulations to model the formation of random copolymers with tunable monomer sequence distributions. Our scheme is based on the original idea proposed a few years ago by Khokhlov and Khalatur [Physica A 249, 253 (1998); Phys. Rev. Lett. 82, 3456 (1999)], who showed that the distribution of species B in A-B random copolymers can be regulated by (a) adjusting the coil size of a homopolymer A and (b) chemically modifying (“coloring”) monomers that reside at (or close to) the periphery of the coil with species B. In contrast to Khokhlov and Khalatur’s work, who modeled the polymer modification by performing the coloring instantaneously, we let the chemical coloring reaction progress over time using computer simulations. We show that similar to Khokhlov and Khalatur’s work, the blockiness (i.e., number of consecutive monomers) of the B species along the A-B copolymer increases with increasing degree of collapse of the parent homopolymer A. A simple analysis of the A-B monomer sequences in the copolymers reveals that monomer sequence distributions in homopolymers “colored” under collapsed conformations possess certain degrees of self-similarity, while there is no correlation found among the monomer sequence distributions formed by coloring homopolymers with expanded conformations. © 2006 American Institute of Physics. [DOI: 10.1063/1.2210011]

I. INTRODUCTION

Previous studies have investigated how monomer sequence distribution affects the solution and interfacial behavior of copolymers.\textsuperscript{1–38} During the past few years, several papers have appeared that focused on designing methods leading to the formation of random copolymers and studying their physicochemical properties. In particular, Khokhlov and co-workers\textsuperscript{17,18,21,22,24,30,32–37} suggested a simple computational scheme for generating copolymers with adjustable monomer sequence distributions, so-called “proteinlike” copolymers (PLCs), via selective chemical “coloring” of parent homopolymers. Khokhlov and co-workers showed that by adjusting the dimension of a parent homopolymer chain (made of, say, A units) and exposing this homopolymer to a coloring species (say, B), a selected number of the A monomers were “colored” with B moieties. Depending on the degree of collapse of the parent homopolymer A coil, different sequence distributions of the B species were realized after such coloring. Thus, while A coils that were fully expanded exhibited a random distribution of the B species after coloring, only those monomers that were present on or close to the periphery of a collapsed coil underwent coloring, which produced random-blocky proteinlike monomer sequences of A and B. The PLCs generated in these simulations were amphiphilic in nature; the colored B units were assumed to be hydrophilic and the original unmodified A units were regarded to be hydrophobic. In a series of papers, Khokhlov and co-workers explored the assembly of such simulated PLCs in the bulk and at interfaces and evaluated various thermodynamic characteristics.\textsuperscript{17,18,21,22,24,30,32–37}

Since the introduction of the coloring scheme of Khokhlov and co-workers, several experimental studies have appeared that aimed at synthesizing random copolymers with tunable monomer sequence distributions. Early works utilized poly (N-isopropyl acrylamide) (PNIPAAm), which undergoes a coil-to-globule transition in aqueous solutions; PNIPAAm expands at temperatures below \( \approx 32\ \text{oC} \) and collapses at temperatures above \( \approx 32\ \text{oC} \). PNIPAAm-based copolymers with adjustable monomer sequence distributions were generated by copolymerizing NIPAam with other monomers \([\text{e.g.,} N\text{-acryloylsuccinimide or glycylidyl methacrylate}^{39–41}]\) (GMA) and vinylpyrrolidone]; additional chemical tailoring was achieved by grafting side chains, e.g., poly (ethylene oxide), onto the GMA groups while at the same time adjusting the number and position of the GMA moieties along the polymer backbone through temperature-driven coil-to-globule transition of PNIPAAm. Other synthetic schemes utilized reactions between hydrophobic and hydrophilic monomers.\textsuperscript{42,44,46} In these schemes the composition and monomer sequence distribution in the final copolymer were varied by adjusting the polymerization conditions through monomer loading, solvent type, and reaction temperature. More recent simulation work performed by Berezkin \textit{et al.} has examined such copolymerization processes.\textsuperscript{38}

Our group has very recently developed another synthetic protocol leading to the formation of PLCs. By using polystyrene (PS) as a parent homopolymer and utilizing electrophilic substitution/addition of bromine in the \textit{para} position of the phenyl ring,\textsuperscript{47} a series of poly(styrene-co-4-
bromostyrene) (PBrS) copolymers with adjustable fractions and sequence distributions of 4-bromostyrene (4-BrS) monomers in PBrS was synthesized in alkyl-halogen solvents, whose theta temperatures (θ) lie in a convenient experimental window (≈6–59 °C). The degree of bromination and the 4-BrS sequencing were adjusted by varying the bromine concentration in the reaction vessel, bromination reaction time, and solvent temperature. A battery of experimental tools, including Kerr effect measurements and interaction chromatography, established the transition from random to solution viscosity and solution coloring reaction in “real time” using computer simulation.

It is generally limited and will take place on a much longer time scale than monomer and coil relaxation times, we perform the coloring reaction in “real time” using computer simulation.

Our computer simulations are carried out in a two-step computational scheme. First, the collapse transition is mapped out using a Monte Carlo simulation with an implicit solvent. These simulations generate the initial homopolymer conformations used during the coloring simulation. The degree, to which the homopolymer coil is collapsed, is achieved by adjusting the inter- and intramolecular potentials acting among the chain monomers. In the second process, another Monte Carlo simulation is conducted, which involves the coloring reaction performed directly on the homopolymer chain of a given state of collapse. The coloring reaction simulation scheme uses an explicit reactant that has a predetermined reaction probability with the monomeric units along the homopolymer. During this simulation a monomer unit along the length of the chain has a certain probability to either react with the reactant or make a conformational move. Both the chain collapse and coloring simulations are performed many times in order to access the statistics of the process. Because polymers undergoing the coil-to-globule transition in solution do not always expose the same monomer units to the periphery of the chain, there is going to be a distribution of the colored blocks in the final PLC after the coloring reaction. We perform a simple analysis that provides insight into the statistical distribution of the monomer sequences among all PLCs created. In particular, we show that while the distribution of the monomer sequences among copolymers created by coloring expanded homopolymers is random, the sequences in PLCs generated by coloring collapsed coils are statistically related.

II. THE MONTE CARLO MODEL

Our Monte Carlo (MC) simulation scheme is based on the bond-fluctuation model (BFM) developed by Carmesin and Kremer. In the BFM, polymers are represented as connected repeat units/monomers residing on a three-dimensional cubic lattice. Each monomer unit is allowed to occupy a single cube and successive monomers along the chain are connected via a predetermined set of bond vectors. The bond vectors are built through all possible permutations and sign inversions of the following vector families: \( P(2,0,0) \cup P(2,1,0) \cup P(2,1,1) \cup P(2,2,1) \cup P(3,0,0) \cup P(3,1,0) \). These vector sets prevent any bond vector from crossing and monomers from overlapping. The chief benefit in using the BFM scheme is that its high coordination number allows one to closely approximate continuum behavior while retaining the advantages of lattice models, such as integer arithmetic and parallelization.

All monomer and polymer moves in our MC simulations follow a single-move MC algorithm, which enables a reasonable acceptance rate when the polymer conformation traverses from the expanded to globular conformations. The algorithm is carried out in the following manner. A single monomer, chosen at random, is translated in a random direction by one lattice spacing. A check is made to see if the resulting move retains the constraints set forth by the BFM. If the restrictions are violated the move is rejected. If the constraints are satisfied, then the energy of the new conformation \( E_N \) is evaluated. The move is accepted with a probability equal to \( \min[1, \exp(-β(E_N-E_0))] \), where \( E_0 \) is the system energy before the move and \( β=1/kT \).

III. MOLECULAR PARAMETERS

A series of simulations was conducted for three chain lengths having degrees of polymerization (DPs) equal to 50, 100, and 150. The simulation cell sizes were kept constant at 25 × 25 × 25, 50 × 50 × 50, and 50 × 50 × 50 for the 50, 100, and 150 length chains, respectively. All simulations were inspected to ensure that chain conformations were not influenced by periodic boundary conditions. In order to simulate the coil-to-globule transition of a single coil in an implicit-solvent manner, we used the potential model developed by Wittkop et al. By using a truncated Lennard-Jones potential, \( U(r) \), within a range of three lattice units and a bond length potential, \( U(l) \), which describes the increasing backbone stiffness with decreasing temperature, the collapse can be simulated without an explicit solvent present. These potentials are expressed as

\[
U(r) = \frac{2ε}{(d-1)^2} \left( \frac{r}{r_0} \right)^{-12} - 2 \left( \frac{r}{r_0} \right)^{-6},
\]

\[
U(l) = \frac{ε}{2} \left[ c_0 + c_1 \left( \frac{l}{a} \right)^2 + c_2 \left( \frac{l}{a} \right)^{12} + c_3 \left( \frac{l}{a} \right)^{18} \right],
\]

where \( k \) is Boltzmann’s constant, \( T \) is the temperature, \( ε \) is an energy unit, and \( d \) is the space dimensionality, \( a \) represents a lattice spacing, \( r \) denotes the distance between monomer units, \( r_0 \) is the equilibrium spacing, and \( l \) is the bond length. The parameters \( c_0 = -207.12, c_1 = 342.88, c_2 = -163.52, \) and \( c_3 = 24.32 \) are taken directly from the report of Wittkop et al. The values of \( U(l) \) and \( U(r) \) as a function of the bond length and lattice separation, respectively, are plotted in...
The simulations commence by placing the polymer chain within the simulation box and equilibrating at high temperature for $50 \times 10^6$ MC steps. The temperature is allowed to decrease slowly at a rate of $4.0 \times 10^{-7} \epsilon/kT$ per MC step; concurrently, the chains are reequilibrated every 62 500 MC steps for $50 \times 10^6$ MC steps. During this re-equilibration process the radius of gyration and system energy are evaluated and recorded every 100 MC steps in order to determine the ensemble averages at the respective temperatures. These simulations allow for a complete mapping of the coil-to-globule transition, which can be physically related to the solvent quality of a real polymer. These expanded and collapsed chain conformations are then used as an input to the coloring simulations, which are described next.

The MC procedure used to simulate the coloring reaction is similar to the one used to mimic the coil-to-globule transition. The cell sizes are kept consistent with the collapse simulations. The MC simulations are initialized using the end configurations of the collapse simulation. Once the chain is placed on the lattice a threefold excess of reactant is added at random to the simulation box. Chain monomers and reactants are chosen at random and are then allowed to either (a) make a conformational/translational move within the lattice or (b) react with a neighboring reactant/monomer. The interactions among chain monomers are still governed by the aforementioned inter- and intramolecular potentials. The reactants are assumed to be neutral to all species within the simulation box. Reactions are allowed to occur with a predetermined probability (set to either 5% or 25%) if the distance between the reactant and monomer along the chain is less than $\sigma$. The simulation progresses until a specified percentage of the chain is colored. In our work, we utilize two different degrees of coloring: 40% and 60%. For each discrete position along the coil-to-globule transition, 100 separate coloring reactions are conducted in order to generate statistical averages.

IV. RESULTS AND DISCUSSION

We start by mapping out the coil-to-globule transition by tuning the inter-and intra-molecular potentials via varying the reduced interaction potential $\epsilon/kT$. In Fig. 2, we plot the homopolymer coil’s radius of gyration ($R_g$) for DP values equal to 50 (■), 100 (●), and 150 (▲) as a function of $\epsilon/kT$. At high temperatures (low $\epsilon/kT$) the coil assumes an expanded state, while at low temperatures (high $\epsilon/kT$) the coil collapses into a globule. While in the expanded state the coil size increases strongly with increasing DP, in the collapsed state varying the DP exhibits only marginal effect on the coil size. Regardless of the chain length, the transition point, defined loosely as an inflection point in the $R_g$ versus $\epsilon/kT$ plot, occurs at $\epsilon/kT \approx 0.225$. The data in Fig. 2 clearly demonstrate that one can simulate the coil-to-globule transition of a single chain using the interaction potentials developed by Wittkop et al.

The polymer conformations generated through the coil-
to-globule transition are then used as inputs for the coloring reaction simulations. The MC simulations were carried out for $\varepsilon/kT$ ranging from 0 (completely expanded coil) to 0.5 (completely collapsed coil). In some situations, obtaining the desired extent of coloring with the completely collapsed configurations proved to be very challenging to achieve in reasonable computational times. In those cases, we set the upper limit for the collapsed chain conformations of the homopolymer at $\varepsilon/kT=0.375$. Figure 3 depicts typical conformations of the PLCs obtained by coloring homopolymers with three different DPs (top panel: DP=50, middle panel: DP=100, and bottom panel: DP=150) at various degrees of collapse ranging from expanded ($\varepsilon/kT=0.0$) to collapsed ($\varepsilon/kT=0.375$) conformations. As apparent from the images, the coil sizes decrease with increasing $\varepsilon/kT$ upon crossing the coil-to-globule transition. The images alone do not provide sufficient information about the distribution of the “uncolored” and colored species along the macromolecule. A simple analysis of the statistical distribution of the monomer sequences in all colored copolymers is presented next.

We first examine the effect of chain conformation before coloring on the blockiness (i.e., number of consecutive monomers) of the colored species in the copolymer by simply plotting the average number of colored sequences as a function of the length of the colored block (i.e., the number of consecutive colored species). The actual number of colored sequences increases with increasing DP. In order to remove the effect of DP, we present the average number of colored sequences by normalizing the actual values per 100 monomers in the macromolecule. In Fig. 4, we plot the average number of the colored sequences as a function of the length of the colored block for chains in the expanded ($\varepsilon/kT=0$), $\theta$ ($\varepsilon/kT=0.225$), and collapsed ($\varepsilon/kT=0.375$) conformations of the parent homopolymer with DPs equal to (a) 50, (b) 100, and (c) 150. The extent of coloring was set to 60%. From the data in Fig. 4, it is apparent that the average number of colored blocks decreases with increasing the length of the colored block. Comparison of the number of colored blocks for the expanded, $\theta$, and collapsed polymer conformations for a given number of colored blocks reveals that the population of short blocks dominates in copolymers colored in the expanded homopolymer state. Concurrently, with increasing the length of the colored block, the number of colored blocks found in polymers colored in collapsed conformations outweighs that of polymers colored in an expanded state. Hence, as expected, the collapsed chain is more susceptible to developing blocks with longer lengths of the colored monomer relative to the chains in expanded conformations.

Our simulations reveal an effect of the DP on the length of the colored blocks. Specifically, an increase in the chain length results in an increase of the fraction of longer colored blocks. For instance, for the chain having a DP=50 the maximum block size is $\approx$12 monomer units, whereas in the macromolecules with DPs of 100 and 150 the maximum block lengths are greater than 15 monomer units. This increase in the number of longer colored blocks is due to the increase in the surface area to volume ratio when considering the collapsed state. Although the DP has increased greatly, the surface area to volume ratio has only increased by 25%–35% from the collapsed to expanded states for the DP=100 and 150 chains, respectively. Therefore, one would expect to see a larger degree of blockiness of the colored moieties in chains with long lengths relative to their respective surface area to volume ratios. While the dependence of the DP on these parameters has been virtually removed by normalizing the number of colored blocks per 100 monomers, it becomes more apparent when examining the average block length and
the number of blocks after the coloring reaction. Figure 5 depicts the average length of the colored blocks as a function of the conformation of the parent homopolymer having DPs equal to (a) 50, (b) 100, and (c) 150. As expected, for each DP the average colored block length increases with increasing the degree of collapse of the homopolymer. In addition, for each polymer conformation the average length of the colored block increases with increasing DP. In the same figure, we also plot the number of the colored blocks in each copolymer as a function of the parent homopolymer conformation. Generally, polymers with a higher DP possess a larger number of colored blocks. As previously, we present the data in a normalized form (per 100 monomers) in order to remove the effect of the DP. The general trend observed reveals that the number of colored blocks decreases with increasing the collapse of the parent homopolymer. The difference between the number of colored blocks detected in the expanded and the collapsed coils is reminiscent of the efficiency of tuning the distribution of the coloring sequences along the macromolecule. From the data shown in Fig. 5, this difference increases with increasing the DP of the parent homopolymer. This behavior results from the fact that polymers with a high DP are capable of forming globulae that are more tightly packed than those formed by collapsing polymers with a lower DP. These findings again illustrate how intimately the degree of blockiness of the colored species is related to the conformational state of the parent homopolymer.

We have also explored the effect of the extent of coloring on the length and distribution of the colored blocks in the copolymer. In Fig. 6, we plot the average length of colored sequences (normalized per 100 monomers) as a function of the length of the colored block in chains in expanded ($e/kT=0$), theta ($e/kT=0.225$), and collapsed ($e/kT=0.375$) conformations of the parent homopolymer with DPs equal to (a) 50 and (b) 100 for a reaction extent of 40%. The trends in the average number of colored blocks as a function of the length of the colored block are similar to those observed for the 60% coloring case. Namely, the average number of the colored blocks decreases with increasing the length of the colored block. In addition, the number of short blocks dominates in copolymers prepared by coloring expanded homopolymers, the number of the long colored blocks increases for copolymers formed from collapsed homopolymers. The main effect of varying the frequency of coloring is exhibited in the total number of the colored blocks and the rate of decrease of the average number of colored blocks with increasing the length of the colored block. A close comparison between the data shown in Figs. 4 and 6 reveals that for short monomer sequences, the average number of colored blocks increases with decreasing DP, while for long monomer sequences, the average number of colored blocks decreases dramatically with decreasing DP.

Decreasing the concentration of the colored species along each macromolecule exhibits a profound effect on the average block length and the number of colored blocks. In Fig. 7, we plot the average length of the colored block as a function of the conformation of the parent homopolymer having DPs equal to (a) 50 and (b) 100. As previously, the extent of coloring is set to 40%. The average block length increases with increasing the degree of collapse of the parent homopolymer. This increase is much smaller than that observed in polymers having 60% of their monomers colored (cf. Fig. 5). The effect of DP on the average block length does not seem to be very significant. In Fig. 7, we also present the number of the colored blocks in each copolymer as a function of the conformation of the parent homopolymer. While the number of the colored blocks is larger than that observed in copolymers having 60% monomers colored (cf. Fig. 5) and decreases with increasing degree of collapse of the parent homopolymer, the latter effects are much weaker relative to those shown in Fig. 5.

In order to gain insight into the statistical distribution of the monomer sequences formed by coloring homopolymers...
under various degrees of collapse, we perform a simple analysis to ascertain the relative sequence matching between the coloring carried out on the expanded and collapsed coils and other copolymers generated from the coil-to-globule transition. The principle of our analysis is illustrated pictorially in Fig. 8. We commence by randomly selecting a copolymer chain produced from the reaction simulations performed using expanded ($\epsilon/kT=0.0$) and collapsed ($\epsilon/kT=0.375$) conformations. The first and last segments of the chosen copolymer are connected thus creating a ring of sequences present in the probe macromolecule. In step 2, we randomly select a monomer along the copolymer chain and select a random sequence length. We refer to this sequence as the probe. In step 3, we pick another copolymer as the target. The target is chosen from any of the copolymers formed from any initial configuration along the collapse transition ($\epsilon/kT$ ranging from 0 to 0.375). Lastly, in step 4 we compare the probe sequence on a one-to-one basis, both in the forward (+) and backward (−) directions, to the target macromolecule in order to determine whether the probe sequence exists anywhere along the length of the target chain. This four-step analysis is carried out repetitively over $1 \times 10^6$ attempts for each $\epsilon/kT$ along the collapse transition in order to generate the matching probabilities between the probe and the target.

In Fig. 9, we plot the match parameter (fraction of matching sequences per trial) for macromolecules colored in the expanded state [Fig. 9(a), $\epsilon/kT=0$] and those colored in the collapsed conformation [Fig. 9(b), $\epsilon/kT=0.375$] for chains having an extent of coloring set to 60%. The data in Fig. 9 reveal that for random monomer sequence distributions ($\epsilon/kT=0$) the probability of matching sequences with all other copolymers generated along the collapse transition...
FIG. 9. Match parameter for chains with degrees of polymerization (DPs) equal to 50 (■), 100 (●), and 150 (▲) as a function of the reduced interaction potential $\varepsilon/kT$ determined for probe macromolecules having a random distribution of “coloring” [(a), $\varepsilon/kT=0$] and “blocky” distribution of “coloring” [(b), $\varepsilon/kT=0.375$].

is approximately equal for a given DP. Importantly, while increasing DP produces a higher match probability, this probability remains independent of the degree of collapse (i.e., $\varepsilon/kT$). In contrast, copolymer sequences generated from a collapsed state have a higher probability of matching chain sequences produced under conditions below the $\theta$ temperature ($\varepsilon/kT>0.2$) relative to those above the $\theta$ temperature ($\varepsilon/kT<0.2$). The effectiveness of matching increases with increasing DP. This increase in sequence matching probability in PLCs relative to the “truly random” copolymers is an inherent property of the PLCs, in which the distribution of the colored sequences bears a memory of the original chain conformation.

As mentioned earlier, one of the parameters we varied in the simulation was the reaction probability between the coloring species and the parent homopolymer. All results presented here have all been obtained with the reaction probability of 25%. We have also carried out similar computer simulations with a lower reaction probability (=5%). Little variation was seen when comparing the 25% and 5% reaction probability data.

V. CONCLUSIONS

We used Monte Carlo simulations with a single-move algorithm and a simulated annealing scheme to map out the coil-to-globule collapse transition of a single chain having degree of polymerization (DPs) equal to 50, 100, and 150. Intermolecular and intramolecular potentials were utilized to simulate this coil-to-globule transition without the use of an explicit solvent. By starting at high temperature and slowly cooling the system down, the coil-to-globule transition was detected at $\varepsilon/kT=0.225$ in all DPs studied. Conformation snapshots generated along the coil-to-globule transition were, in turn, used as inputs for “coloring” reactions performed directly on the homopolymer. In our simulation scheme, the parent homopolymer with an adjusted conformation was subjected to a threefold excess of reactant until a predetermined extent of coloring (40% or 60%) was achieved. The simulation results revealed that regardless of the DP of the parent homopolymer, all monomers along the fully expanded chains possessed an equal probability of coloring. In contrast, the probability of coloring monomer units in collapsed chains was distributed widely. Increasing the chain length resulted in monomer sequences with broader distributions of colored units. Furthermore, expanded coils exhibited an upsurge in the frequency of occurrence of short sequence lengths of the colored units, while collapsed coils contained larger fractions of long sequential monomer lengths. A higher frequency of long colored blocks was also detected upon increasing the chain length of the parent homopolymer. A simple statistical analysis revealed that while the sequence distribution of the colored species in polymers colored in an expanded state was random, homopolymers colored under collapsed conformations possessed statistically self-similar sequence distributions. The latter observation and the fact that distinct coloring sequences can be achieved by coloring parent homopolymers over time (rather than performing instantaneous coloring as performed by Khokhlov and co-workers) have important implications for preparation of random copolymers with distinct and adjustable monomer sequences. Recent work performed in our laboratory indicates that experimental realization of such coloring reactions is indeed possible and that the random copolymers thus created exhibit very distinctive interfacial and bulk characteristics.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation through the Grant No. DMR-0353102.


Strictly speaking, because BFM is a coarse-grained model, the simulation monomers do not correspond to monomers present in a real chain.