HA (Hydrophobic/Amphiphilic) Copolymer Model: Coil−Globule Transition versus Aggregation

Valentina V. Vasilevskaya,[†] Alexei A. Klochkov,[†] Alexei A. Lazutin,[‡] Pavel G. Khalatur,[*,§] and Alexei R. Khokhlov[‡,§]

Institute of Organoelement Compounds, Russian Academy of Sciences, Moscow 117823, Russia; Physics Department, Moscow State University, Moscow 119899, Russia; and Department of Polymer Science, University of Ulm, Ulm D-89069, Germany

Received December 22, 2003; Revised Manuscript Received May 10, 2004

ABSTRACT: For simulating hydrophobic−amphiphilic (HA) copolymers, we have developed a "side-chain" HA model in which hydrophilic (P) interaction sites are attached to hydrophobic (H) main chain, thereby forming amphiphilic (A) monomer units, each with dualistic (hydrophobic/hydrophilic) properties. Using this coarse-grained model, we performed molecular dynamics simulations of the hydrophobically driven self-assembly in a selective solvent, for both single-chain and multichain systems. The focus is on the regime in which H and P interaction sites are strongly segregated. Single-chain simulations are performed for copolymers with the same HA composition but with different distribution of H and A monomer units along the hydrophobic backbone, including regular copolymers comprising H and A units in alternating sequence, $(HA)_x$, regular multiblock copolymers $(H_LA_L)_x$ composed of H and A blocks of equal lengths $L = 3$, and quasi-random proteinlike copolymers having quenched primary structure. In a solvent selectively poor for H sites, the proteinlike polyamphiphiles can readily adopt spherical-shaped compact conformations with the hydrophobic chain sections clustered at the globular core and the hydrophilic groups forming the envelope of this core and buffering it from solvent. Because of the fact that these globules are size- and shape-persistent objects, they maintain their morphological integrity even in rather concentrated solutions where no large-scale aggregation is observed. Moreover, we find that the population of aggregates generally decreases with worsening solvent quality. The compact conformations of long regular copolymers tend to be strongly elongated in one direction.

I. Introduction

One of the most fascinating features of polymer systems relates to the fact that a phase transition can occur at the level of a single chain. The best-known example is the coil-to-globule transition. The coil collapse problem is of interest not only because of its fundamental importance as related to polymer nanostructures but also because of various biological applications such as protein folding. Though the theory of this transition is well understood,[1,2] related experimental studies face difficult challenges: the coil-to-globule transition occurs at poor solvent conditions and is almost invariably accompanied by precipitation. This significantly complicates the data analysis, preventing the clear correlation between the experimental observations and the molecular parameters. At the same time, it is well-known that globular proteins differ from the globules of synthetic water-insoluble polymers in three main ways. (i) Globules formed from homopolymers and random copolymers are typically insoluble in aqueous medium, whereas many of globular proteins are water-soluble. (ii) In a poor solvent, when polymer segments attract each other strongly, synthetic globules stick together and form intermolecular clusters or aggregates even in very dilute solutions. It is the aggregation that makes them insoluble. At sufficient concentration, aggregation results in polymer precipitation. Globular proteins fold and may form small aggregates (quaternary structure) preserving their globular native state, but even these are still water-soluble. (iii) Globular proteins are mobile in solution due to their solubility. The aggregation of synthetic globules dramatically slows down their diffusion in the polymer precipitate.

It is thought that protein globules are soluble in water because of the special primary sequence: hydrophobic amino acids effectively join together, and by doing so, they are avoiding the surrounding aqueous environment. One of the major driving forces in the folding of proteins is to place each of the types of amino acids (hydrophilic and hydrophobic) in an environment appropriate for its solution properties. This can be achieved by locating the majority of hydrophilic amino acids on the globule exterior where they can bond to water molecules, while most hydrophobic amino acids are clustered in a central core where they bind to each other in an effectively water-free environment.

Having in mind this fundamental principle of protein organization, we can now discuss how to optimize copolymer sequences to obtain better aggregation stability. Since globule formation and solution stability cannot be simultaneously realized with any random sequence of monomer units, we can formulate the following problem: is it possible to design such a sequence of an HP copolymer (i.e., a copolymer consisting of segments of two types, H and P) that in the dense globular conformation all the hydrophobic H segments are in the core of this globule while hydrophilic P segments form the envelope of this core? Indeed, it is clear that a simple way to protect designed copolymer chains from aggregation would be to make the chains compact with a stable core−shell structure.

[†] Russian Academy of Sciences.
[‡] Moscow State University.
[§] University of Ulm.
* To whom correspondence should be addressed. E-mail: khalatur@germany.ru.

10.1021/ma0359741 CCC: $27.50
Published on Web 06/18/2004

In the past few years, a number of groups have analyzed the aggregation mechanism of heteropolymers using computer simulations employing both lattice and off-lattice representation based on the so-called HP model first introduced by Lau and Dill.[3,4] This two-letter model of a linear hydrophobic/hydrophilic macromolecule involves only two types of polymer segments, H and P, and reflects the spirit of minimalist models, in that it is simple yet based on a physical principle. With this model, Abkevich et al.[5] proposed a simple algorithm that biases sequence sampling toward compact and water-soluble sequences. Using Monte Carlo simulations, the competition between chain folding and aggregation was studied by following the simultaneous folding of two designed copolymer chains within the framework of a lattice model.[6] It was found that aggregation is determined by partially folded intermediates formed at an early stage in the folding process. Giugliarelli et al. investigated how the interaction potentials affect the solubility and compactness of short heteropolymers on a two-dimensional lattice.[7] Maximally compact conformations were found to be destabilized in solution as the intermolecular potential varies. Timoshenko and Kuznetsov[8] simulated the formation of clusters consisting of several linear heteropolymers in dilute solutions. They found that, at relatively low concentrations of heteropolymers with sufficiently strong competing interactions, such clusters ("mesoglobules") are more stable in a selective solvent as compared to single chains. Bratko and Blanch[9,10] considered a lattice model designed to examine the competition between intramolecular interactions and intermolecular association, resulting in the formation of aggregates of misfolded chains (see also refs 11 and 12). The question formulated in the previous paragraph was directly addressed in refs 13 and 14. The biomimetic approach proposed in these papers includes the computer generation of a special primary sequence in two-letter copolymers from some specific conformation of a homopolymer chain ("parent" conformation) via a one-step "coloring" procedure of the instant image of such chain. In refs 13 and 14 this conformation-dependent sequence design was first realized for the case when the "parent" conformation corresponds to a dense globule of a homopolymer chain: the monomer units lying in the core of such a globule were designated as H units and called hydrophobic, while the monomer units belonging to the surface of the "instant image" of a globule were designated as P units and called hydrophilic. The corresponding polymers having a core−shell microstructure were called "proteinlike" copolymers. The motivation behind this design procedure was that each protein globule has a hydrophobic core and a hydrophilic envelope, so when two globules meet, they only contact with their hydrophilic peripheral parts, which are obviously not sticky in water. Later, proteinlike copolymers were synthesized in real chemical experiments[15−17] where the role of "coloring" is played by the chemical reaction of chain segments with a reagent, which converts solvophobic groups to solvophilic ones, thereby screening part of the solvophobic core from the aqueous solution. Although such copolymers are not exhibiting so high a tendency to aggregation as their random and random-block counterparts,[17] nevertheless, they are not completely protected from aggregation,[15,16] in contrast to many real protein globules. Therefore, it is instructive to look at other factors that can be responsible for aggregation stability.

Real proteins are built up from both hydrophobic and polar amino acid residues; some of the latter can be charged. Many of the conformational and collective properties of proteins are due to a complex interplay between short-range (hydrophobic) effects and long-range (Coulomb) interactions. Electrostatic effects also can determine some of the unique solution properties of globular proteins. Recently, we have reported the results of molecular dynamics simulations of charged proteinlike hydrophobic−hydrophilic copolymers with a fixed charge distribution at poor solvent conditions.[18] Indeed, for this model we have observed a solution of nonaggregating polymer globules which form a stable array (due to a large net charge of each chain) resembling a charge-stabilized (micro)suspension or microgel phase. A solution of globules has also been seen in the molecular dynamics simulations of Kremer and coworkers,[19,20] who studied highly charged regular copolymers in a poor solvent. It should be kept in mind, however, that many of the nonaggregating proteins are uncharged.

Another way leading to nonaggregating copolymers may be connected with the molecular design of their monomer units. It is well-known that in the large majority of real water-soluble polymers many monomer units have a *dualistic* (hydrophobic/hydrophilic) character, that is, repeating polymer unit, which is usually considered as pure hydrophobic or pure hydrophilic, actually incorporates both hydrophilic and hydrophobic parts concurrently. Many of amino acids also contain both hydrophilic and hydrophobic groups simultaneously, and strictly speaking, the interaction between such amino acid residues in proteins cannot be literally reduced to pure hydrophilic or pure hydrophobic interactions, as it is presupposed in the standard HP model.[3,4] In this model, an amphiphilic polymer is considered as a chain in which H and P groups are assumed to be pointlike interaction sites distributed along the chain in a linear fashion. The "beads-on-a-string" HP model is very simple and computationally efficient, but its principal disadvantage is the representation of each monomer unit of an amphiphilic chain as a pointlike interaction site.

To eliminate this drawback, we have recently introduced an extended variant of the HP model that explicitly takes into account the amphiphilic nature of hydrophilic segments.[21] This off-lattice model generalizes the HP model of Lau and Dill[3] by representing the hydrophobic backbone and hydrophilic side groups of a polyamphiphile with tangent spheres. With this representation, each amphiphilic monomer unit (A) can be treated as hydrophobic/hydrophilic HP "dumbbell". In our previous work,[21] using this "side-chain" model, we have studied homopolymers consisting of amphiphilic monomer units (poly-A). Our simulations have shown that such a trivial modification of the standard HP model can lead to some nontrivial consequences when studying the collapse of single-chain amphiphilic homopolymers and that for this model a variety of novel structures with high complexity are possible in a poor solvent, depending on the interaction between hydrophobic (H) and hydrophilic (P) sites. Specifically, the thermodynamically stable anisometric structures have been observed, including disklike structures, stretched-necklace-like conformations, and cylindrical-shaped conformations. It is important to note that these compact

microstructures are formed due to strong intramolecular segregation of chemically different H and P groups tending to minimize the number of H−P contacts unfavorable under poor solvent conditions for H sites. As a result, a core composed of mainly hydrophobic groups turns out to be surrounded by a thin dense "skin" composed of mainly hydrophilic groups. Generally speaking, the same structural organization is characteristic for globular proteins, allowing them to dissolve in water and preventing their solution aggregation. Therefore, it is believed that, following this line, we can construct model copolymers leading to the formation of nonaggregating heteropolymer globules, with the ultimate objective of learning how to manipulate the polymer chemistry and system conditions in order to preclude the aggregation processes.

In this paper we extend our previous consideration[21] for the case of copolymers containing both pure hydrophobic (H) and amphiphilic (A) groups. We will call such copolymers "HA copolymers". Using this model, we will simulate both regular HA copolymers and quasi-random HA copolymers having proteinlike sequences.[13] In particular, our aim is to study the conformation of single chains as well as their aggregation under poor solvent conditions for hydrophobic groups. To this end, multichain systems in a selective solvent are simulated. The solution morphologies and properties of these systems depend on the amphiphilic primary structure, temperature, and concentration and on the interactions between the system components.

In the next section, we describe the details of the model and simulation. Then we present simulation results for isolated model polyamphiphiles and for systems containing 27 polymers per simulation box. Results of these calculations provide insights into the morphologies and energetics of interchain association. In the final section, our findings are summarized and directions of future work are outlined.

II. Model and Simulation Technique

A schematic representation of the model copolymer used in our study is shown in Figure 1. In the numerical simulations, we employ a continuum space (bead−rod) model, as opposed to widely used lattice models, since the latter have intrinsically discretized dynamics of a rather arbitrary nature and have slow relaxation for a dense globular state. Each hydrophobic monomer unit (H) is considered as a single interaction site (bead); each amphiphilic group (A) is modeled by a "dumbbell" consisting of H and P beads linked by rigid bonds of a fixed length. The hydrophobic units are connected with each other and with H site of amphiphilic groups in a linear fashion and form the backbone of the copolymer. The HA composition is fixed at about 1:1. Namely, a copolymer having the hydrophobic backbone of length N includes also $(N + 1)/2$ hydrophilic beads attached to the backbone; the total number of beads is equal to $\mathcal{N} = (3N + 1)/2$. We simulate copolymers with different distribution of H and A monomer units along the main chain. Namely, we study (i) regular copolymers comprising H and A units in alternating sequence, $(HA)_x$, (ii) regular multiblock copolymers composed of H and A blocks of equal lengths, $(H_L A_L)_x$, with $L = 3$, and (iii) the quasi-random proteinlike copolymers with quenched primary structure.[13]

The time evolution of the system is determined by Newton's equations that are solved by using the method

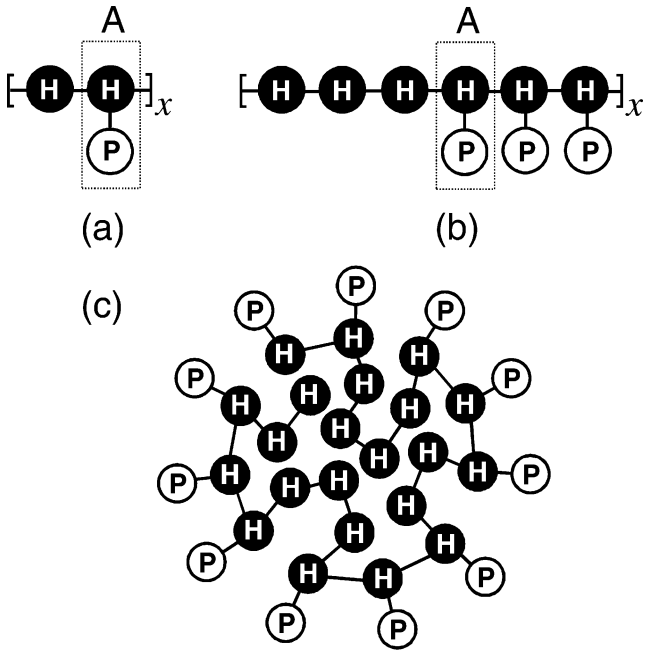

Figure 1. HA "side-chain" models of amphiphilic copolymers: (a) regular alternating copolymer, (b) regular multiblock copolymer, (c) proteinlike copolymer. Each hydrophobic monomer unit (H) is considered as a single interaction site (bead); each amphiphilic group (A) is modeled by a "dumbbell" consisting of hydrophobic (H) and hydrophilic (P) beads.

of molecular dynamics (MD).[22] In MD simulations, the bonds within a given macromolecule are constrained to length $b = 1.0$ using the RATTLE algorithm.[23]

Excluded volume between any nonbonded particles is included via a repulsive Lennard-Jones (LJ) potential. This repulsive, or reference, interaction is defined by a decomposition of the full LJ 6−12 potential as proposed by Weeks, Chandler, and Andersen (WCA)[24]

$$u_{ev}(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} + \frac{1}{4}\right], \quad r_{ij} \le r_0$$
$$= 0, \quad r_{ij} > r_0 \qquad (1)$$

where r is the distance between the two interacting particles and $r_0 = 2^{1/6}\sigma$ is a cutoff distance. The parameter ϵ governs the strength of interaction and controls an energy scale, whereas σ defines a length scale. We assume that $\sigma = \epsilon = 1$ for all interactions. All of our results are therefore reported in terms of these natural units.

In addition to the excluded-volume potential (1), the nonbonded beads interact via a Yukawa-type potential, for which we use the following form:

$$u_s(r_{ij}) = \frac{\epsilon_{\alpha\beta}\sigma}{r_{ij}} f(r_{ij},r_c)h(r_c - r_{ij}) \qquad (2)$$

Here, $f(r_{ij},r_c) = [1 - (r_{ij}/r_c)^2]^2$ is the screening function, $h(r)$ is the Heaviside function, the parameter $\epsilon_{\alpha\beta}$ ($= \epsilon_{HH}$, ϵ_{PP}, ϵ_{HP}) sets the amplitude of the nonlocal site−site interactions, and r_c is the screening length, i.e., the cutoff distance for these interactions. In the model, this potential describes the solvent-mediated short-range hydrophobic−hydrophilic interactions. The characteristic energies of H−H, H−P, and P−P interactions are considered to be variable parameters (all the energies are measured in units of ϵ). For $\epsilon_{\alpha\beta} = 0$, there is no

additional repulsion (attraction) between chain beads except that corresponding to the excluded-volume potential (1). The (negative) parameter ϵ_{HH} is chosen to promote the collapse of the chains. The nonzero (positive) parameters ϵ_{HP} and ϵ_{PP} characterize the H−P and P−P repulsion and promote the intrachain segregation of H and P beads. For simplicity, we assume that $\epsilon_{HP} = \epsilon_{PP} = -\epsilon_{HH} \equiv \psi$, where ψ denotes the strength of the interactions between beads. With these energy parameters, we simulate the chains in a selective solvent, which is a poor solvent for H sites and a good solvent for P sites.

Explicitly, no solvent particles are included in the simulations, as we wish to understand the properties of the system determined solely by the polymer; that is, the solvent is represented by a 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 the frictional term and the Langevin uncorrelated noise term $\mathbf{R}_i$ which is connected with the viscosity of the solvent Γ through the fluctuation−dissipation theorem, $\langle R_{\alpha i}(0) \cdot R_{\alpha i}(t)\rangle = 2\Gamma_i k_B T \delta(t)$, where $\alpha = x, y, z$; $i = 1, 2, ..., nN$; T is the reference temperature; and n is the number of chains in the system. This term ensures that the temperature is kept constant.[22] We take the parameter Γ 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 S_i for each specified bead.[25] Having S_i, one can define Γ_i as $\Gamma_i = \Gamma_0 S_i/S_{max}$, where S_{max} is the maximum solvent-accessible surface area of a bead for the model under study and the reference value of Γ_0 is taken to be equal to unity (for more details, see ref 26).

To keep close contact to the traditional treatment of polymer solutions used in theory, it is instructive to define the following partial bare second virial coefficients as an integral measure of solvent quality

$$v_{\alpha\beta}^{(2)}(\epsilon_{\alpha\beta}) = 2\pi\int_0^{r_c}\{1 - \exp[-u_{\alpha\beta}(r)/k_B T]\} r^2\, dr \quad (3)$$

where $u_{\alpha\beta}(r) = u_{ev}(r) + u_s(r)$ and α, β = H, P. Using $v_{\alpha\beta}^{(2)}(\epsilon_{\alpha\beta})$, we define the normalized value $v_{\alpha\beta} = {}^1/_4\, v_{\alpha\beta}^{(2)}(\epsilon_{\alpha\beta})/v_{\alpha\beta}^{(2)}(0)$ and then introduce the linear combination

$$\tilde{\chi} = v_{HP} - (v_{HH} + v_{PP})/2 \quad (4)$$

In principle, this parameter represents the kind of interactions used in simple lattice systems such as in Flory−Huggins theory.[27] Unlike the standard Berthelot rule for energy parameters $\epsilon_{\alpha\beta}$,[22] the combination (4) allows for the introduction of an arbitrary extra interaction between species of α and β types while keeping the same interaction between like species. The parameter $\tilde{\chi}$ is similar to the Flory−Huggins interaction parameter χ and characterizes solvent quality in an integral manner.[28] In particular, for an athermal good solvent ($\epsilon > 0$, $\psi = 0$) we have $v_{\alpha\beta} = {}^1/_4$ and $\tilde{\chi} = 0$. This state can be considered as a reference state. When $\epsilon = \psi = 0$, we deal with a phantom chain. At a fixed $\psi > 0$ and $\epsilon > 0$, the interaction parameter $\tilde{\chi}$ increases as temperature is reduced. In reality, when the temperature is fixed, the change in $\tilde{\chi}$ can be due to variation of the solvent composition, e.g., concentration of some denaturation agent. In principle, for a given solvent and temperature, the value of $\tilde{\chi}$ can be obtained from thermal effect of transfer experiments. For the model studied here, sufficiently large values of $\tilde{\chi}$ ($\tilde{\chi} \gtrsim 1$) correspond to a poor solvent (see below). Solvent quality becomes poorer with decreasing T or with increasing $\tilde{\chi}$.

To obtain single proteinlike HA copolymers, the following computational strategy was employed:

(i) A self-avoiding homopolymer chain of a given length N is randomly generated.

(ii) The collapse of the chain is performed at $\epsilon_{HH} = -3$ and $T = 1$. Then this conformation is equilibrated for 5×10^5 integration time steps, which is more than enough for the chain to become a globule. The resulting homopolymer globule was molten; i.e., the temperature was not too low to cause rigid freezing.

(iii) The surface of the globule is experiencing a chemical reaction ("polymer-analogous transformation"); this means that the monomer units at the surface, instead of being hydrophobic, are acquiring amphiphilic properties: H + P → A. The process is modeled as a step-by-step chemical reaction of the attachment of P monomers to the hydrophobic backbone. Our algorithm simulates a reaction in a $3d$ cube utilizing periodic boundary conditions. Initially, a homopolymer hydrophobic globule and N free hydrophilic monomers were placed in a cube. Beads H and P are assumed to be monofunctional, and their orientation is not distinguished. If a freely diffusing P monomer approaches within the prescribed distance R_{min} ("reaction radius") to an H bead on the chain, a bond can be formed between the two. The reaction radius was set to $R_{min} = b \pm \delta$, where $\delta = 0.01\sigma$; that is, a hydrophilic monomer was allowed to be grafted to the chain when the corresponding distance r was in the range $b - \delta \le r \le b + \delta$. The reaction is considered as a sequence of the alternating steps: grafting of a new P monomer to the chain and the subsequent relaxation of the chain. Note that the reaction time τ_R that is allowed for the annealing between successive P monomer graftings can be used as a control parameter to produce characteristically different chain conformations. The process is terminated when the required number of the hydrophobic monomer units is transformed into A type. Thus, the composition of the resulting copolymer is constrained so that there are N hydrophobic and $(N + 1)/2$ amphiphilic monomer units, and after that the primary structure of the copolymer is frozen. As a result, we obtain a "parent" proteinlike amphiphilic globule having a core−shell structure.

The primary structure of an amphiphilic proteinlike copolymer can be characterized by its composition, by the average lengths of the hydrophobic and amphiphilic blocks, L_H and L_A, and by the specific distribution of H and A units along the chain. For example, for the 127-unit proteinlike copolymer generated in this work, we find $L_H \approx L_A = L = 2.8$. By definition, for alternating and multiblock copolymers studied here, one has $L = 1$ and $L = 3$, respectively.

To obtain a multichain system, the parent globules of the copolymers were replicated 27 times. The identical images of the parent globules were initially placed on a regular simple cubic lattice spanning a periodic MD box without overlap between them and then were randomly reoriented. Then, the system constructed in this way was relaxed for a long time. At this initial stage, velocities were rescaled at each time step to compensate for the large amount of heat produced in the system.

The equations of motion in conjunction with the constraints of fixed bond lengths were solved iteratively using a Newtonian iteration procedure with the time step $\Delta t = 0.01\sigma\sqrt{m/\epsilon}$, where $m = 1$ is the mass of chain bead. For more details on the integration algorithm used in this study and the bond-constraint method, we refer to our previous publications.[18,29] The nonbonded interactions were computed using the method of lights to locate neighboring particles.[30] In the following, the energy parameter $\epsilon_{\alpha\beta}$ is fixed and the reference temperature T is varied in the range from 0.5 to 20 ϵ/k_B. This corresponds to the range of the interaction parameter $\tilde{\chi}$ from 0.38 to 23.0. As has been mentioned above, for the model under study, the variation in T or $\tilde{\chi}$ is directly connected with the variation in solvent quality. The variation of the interaction parameter $\tilde{\chi}$ allowed covering the wide range of solvent quality, from a good solvent, with dominating excluded-volume effects at small $\tilde{\chi}$, to a poor solvent, with collapsed chains at large $\tilde{\chi}$. It should be stressed that a variation of $\tilde{\chi}$ in our simulations changes simultaneously both the strength of repulsive hydrophilic interactions and the intensity of short-range attractive hydrophobic interactions. This means that, similar to real experiments, a change of $\tilde{\chi}$ influences all relevant quantities of the system in a rather complex manner.

In the case of single-chain systems, we simulated a range of chain lengths up to $N = 255$ for proteinlike copolymers and up to $N = 1023$ for regular copolymers. For the mutichain simulations, chains with relatively short backbone length, $N = 127$, were used. The cubic computational box with periodic boundary conditions contained $n = 27$ copolymers (the total number of beads $n\mathcal{N} = 5157$). For more accurate results, it may be necessary to model a larger number of chains; however, because of the computational difficulties this would involve, we currently restricted ourselves to 27 chains. The entire system of n chains was enclosed in a periodic cubic box with side length L_{box} ($L_{box} = 25\sigma$; volume $V = L_{box}^3$). The number density of the chain beads is $\rho = n\mathcal{N}/V = 0.33\sigma^{-3}$. (Here, "density" is a synonym of "concentration".) The value of ρ is much higher than the overlap concentration of the coils. Note that the classical chain overlap concentration, $\rho^* \propto \mathcal{N}/(^4/_3\pi R_g^3)$, where R_g is the average radius of gyration of the chain in a good solvent, is found to be $6 \times 10^{-2}\sigma^{-3}$ for the model under study. The initial systems were equilibrated for about 2×10^6 time steps, and then the production runs were performed. Equilibration was checked via the observation that the same results can be obtained by starting from different initial configurations and by running the simulations for more time steps. Typically, each production run was from 2×10^6 to 5×10^7 time steps, depending on the chain length and temperature.

III. Results and Discussion

Although we are primarily interested in the intermolecular effects, we first characterize the behavior of single amphiphilic copolymers that we use as a reference for comparison with multichain systems. In particular, we start our discussion by estimating the reference temperature at which the collapse transition takes place in the copolymers for the low-concentration limit. This problem is a first step toward a more precise understanding of collective properties. Also, we analyze in details the equilibrium size and shape of compact conformations.

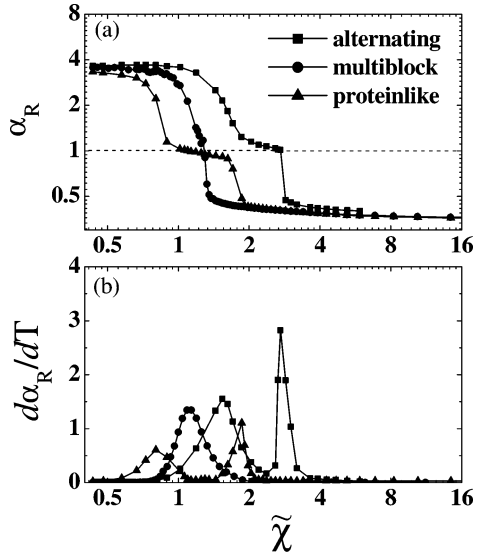

Figure 2. (a) Expansion factor, $\alpha_R = R_g^2/R_{0g}^2$, as a function of the quality of the solvent, $\tilde{\chi}$, in a log−log scale for (■) alternating, (●) multiblock, and (▲) proteinlike copolymers with hydrophobic chain length $N = 127$. Dashed line shows the value of α_R for a phantom chain. (b) Derivatives $\partial\alpha_R(T)/\partial T$ calculated after smoothing of the simulation data and subsequent numerical differentiation plotted vs $\tilde{\chi}$ in a semilogarithmic scale.

1. Single-Chain Conformation and Collapse Transition. a. Gyration Radii. To obtain a quantitative estimate of the polymer size, we calculated the mean-square radius of gyration R_g^2 of individual chains. The simulations at low temperature, or equivalently at large interaction parameter $\tilde{\chi}$, were repeated several times, starting from different initial states. For each polymer chain, we also found mean-square radius of gyration, R_{0g}^2, for a phantom (unperturbed) chain of the same length.[31] Having the values of R_{0g}^2, one can define the following expansion factor $\alpha_R = R_g^2/R_{0g}^2$. This quantity is shown in Figure 2a as a function of the interaction parameter $\tilde{\chi}$ for 127-unit amphiphilic copolymers with different primary structure. It is seen that the chain size decreases with increasing $\tilde{\chi}$, as expected. At small values of $\tilde{\chi}$ (or at high temperature), chain contraction driven by attractive H−H interactions is mainly resisted by the configurational entropy, whereas in the collapsed state repulsive interactions are of paramount importance.

At sufficiently large values of $\tilde{\chi} = \tilde{\chi}_0$, depending on primary structure, the state where $R_g^2 = R_{0g}^2$ (i.e., $\alpha_R = 1$) is reached. We see that for proteinlike copolymer this occurs at smaller $\tilde{\chi}$ than for alternating and multiblock copolymers with the same HA composition. This tendency holds for all the chain lengths studied here. The alternating polyamphiphile demonstrates the highest value of $\tilde{\chi}_0$. For relatively short chains studied in this work, the value $\tilde{\chi}_0$ was found to be dependent upon chain length N. In the case of proteinlike copolymers, we found that the $\tilde{\chi}_0$ value decreases from 1.58 to 1.09 when N is increased from 31 to 255. Alternating copolymers behave otherwise: the $\tilde{\chi}_0$ value increases from 1.61 to 2.71 when N varies from 31 to 255.

Note that the average block length of 127-unit proteinlike sequence is almost the same as for the multiblock sequence, $L \approx 3$. However, the values of $\tilde{\chi}_0$ found for these two copolymers are noticeably different: $\tilde{\chi}_0 = 1.11$ and $\tilde{\chi}_0 = 1.29$ for proteinlike and multiblock copolymers, respectively (see Figure 2a). Therefore, one

can conclude that the properties of proteinlike copolymers are determined mainly by the specific design of their sequence rather than by the average blockiness of the sequence. This can be explained by the presence of the long-range correlations in the primary sequences of proteinlike copolymers. As has been shown, these correlations obey the so-called Levy flight statistics with a high dispersion and a slow (exponential) decrease in the block length probability.[32] There are two physical factors that influence the chain contraction of random (quasi-random) copolymers: the average block length, L, and the dispersion of L. It is known that an increase in L promotes the transition from swollen to contracted state while a larger dispersion leads to an opposite effect.[33] For a sufficiently long proteinlike sequence, very long blocks can be found in it, and it is this factor that plays a dominant role. As the total chain length increases, the difference in the behavior of regular copolymers, having a fixed L, and proteinlike copolymers, for which L gets larger, becomes more and more pronounced. Even for relatively low fraction of long blocks in the chain, their effect can be decisive. This is precisely the major reason behind the specific behavior of proteinlike copolymers, including the way of self-organization.

At a larger undercooling, after reaching the $\alpha_R = 1$ state, chains may achieve some partially ordered conformations. We stress here that in our simulation an ordered conformation should not be viewed as a unique state, but as an average over (closely related) conformations whose number is still very large, though strongly reduced in comparison with a coillike state.

Figure 2a shows that the chains undergo a transition to a compact (folded) state from an ensemble of unfolded configurations. For alternating and proteinlike copolymers with $N = 127$ the chain contraction is a two-stage process, while a multiblock amphiphilic chain of the same length undergoes a single continuous transition. The reason for a two-stage transition for alternating and proteinlike copolymers observed in Figure 2a will be explained later. Here we would like to emphasize the following two points connected with this result. (i) The number of steps (stages) in the collapse transition depends on the chain length N. The shorter is the chain, the smaller is this number. Our data (not presented here) show that short enough chains always exhibit a single-stage collapse transition and that the number of steps is increasing with the increase of N. (ii) For a multistage process, only the transition related to the highest value of $\tilde{\chi}$ leads to the formation of the final globule. The corresponding transition temperature T^*, or interaction parameter $\tilde{\chi}^*$, can be estimated from the inflection point on the $\alpha_R(T)$ curve.

The derivatives $\partial\alpha_R(T)/\partial T$ calculated after smoothing of the simulation data and subsequent numerical differentiation are plotted versus $\tilde{\chi}$ in Figure 2b. They exhibit pronounced peaks, which signal a transition. We see that for alternating and proteinlike copolymers there are two peaks reflecting a two-stage scenario of the chain contraction, while one peak is observed for multiblock copolymer. The wider peaks in the region of small $\tilde{\chi}$ can be attributed to the initial contraction of the swollen chain while the final transition to a collapsed state with $\alpha_R < 1$ is associated with the more narrow peaks at larger $\tilde{\chi}$. It is natural to consider the $\tilde{\chi}^*$ value, corresponding to these low-temperature peaks, as a critical parameter at which the final contraction of the chain takes place. Note that in order to estimate $\tilde{\chi}^*$ or T^*, it is easier to use the radius of gyration (or α_R) than the specific heat. At $\tilde{\chi} < \tilde{\chi}^*$, the polymer is in somewhat expanded conformation and collapses when $\tilde{\chi}$ becomes larger than $\tilde{\chi}^*$, indicating a final contraction of the chain. In the range $\tilde{\chi} > \tilde{\chi}^*$, an increase in $\tilde{\chi}$ results in a consistent decrease in α_R, suggesting that the size of the polymer decreases progressively toward more compact and organized structures. We can refer to this regime as the molten globule state.[34,35]

The reason for multistage coil−globule transition which we observed for some of the sequences and chain lengths can be explained by specific contraction mechanism first described for amphiphilic homopolymers in our previous article.[21] The conformational transitions characteristic of such polymers resemble in some aspects the coil-to-helix transition[36] or the so-called zipping transition.[37] Both these mechanisms arise mainly from strong attractive interactions among near neighbors along the chain, resulting in the local chain crumpling. In other words, chain contraction arises from elementary structures that are controlled by local contacts, which eventually build the hydrophobic nucleus, while contacts between the sections distant along the chain are practically not possible. As a result, the collapse of sufficiently long chains first occurs as intrachain self-organization on a local scale. Therefore, this is not a true coil-to-globule transition[1] and actually not a phase transition at all. It is clear that this specific behavior is due to strong H−P repulsion that leads to the resulting microstructures with strongly segregated hydrophilic and hydrophobic regions. Generally speaking, the multistage contraction observed in Figure 2 can be explained as the formation of initial compact microdomains, which gradually increase in size and merge with each other before reaching a point $\tilde{\chi}^*$ corresponding to transition into final globular structure.

b. Hydrophobic Clusters. To analyze in more details the process of intrachain self-organization observed for HA copolymers, we calculated the $\tilde{\chi}$ dependence of the average size of hydrophobic clusters.

To determine whether a particle in our system is free or clustered, we need to define what is a cluster. In the present study, this is a geometric notion. We identify two hydrophobic beads belonging to a cluster when they are not connected by "chemical bond" and the distance separating them is $r \le r_0$ (see eq 1). Any bead not belonging to a cluster is defined as "free" bead. With this definition, we found the average hydrophobic cluster size $\langle m \rangle$.

Figure 3 presents the ratio $\langle m \rangle/N$ as a function of $\tilde{\chi}$ for 127-unit HA copolymers. For all the copolymers, we have $\langle m \rangle \approx 1$ in the high-temperature region (at $\tilde{\chi} \lesssim {}^1/_2$). At sufficiently large $\tilde{\chi}$, practically all H beads are assembled into a single cluster with $\langle m \rangle/N \approx 1$. Heating leads to the destruction of this cluster. In the case of alternating and proteinlike chains, the cluster disintegrates in a jump-wise manner slightly below the transition value $\tilde{\chi}^*$. This results in the formation of hydrophobic "splinters" of smaller size. It is interesting to note that for 127-unit alternating copolymer these "splinters" have practically equal size, i.e., $\langle m \rangle/N \approx {}^1/_2$. For proteinlike copolymer, however, the hydrophobic beads are distributed between hydrophobic domains in a somewhat different way. From the analysis of the cluster size distribution function (not shown) we found that in this case there is a single large "splinter" and many small

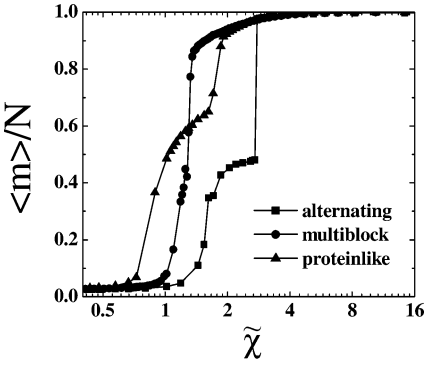

Figure 3. Ratio $\langle m\rangle/N$ ($\langle m\rangle$ being the average hydrophobic cluster size) as a function of the quality of the solvent, $\tilde{\chi}$, in a semilogarithmic scale for (■) alternating, (●) multiblock, and (▲) proteinlike copolymers with hydrophobic chain length $N = 127$.

clusters as well as free H beads. For a regular multiblock copolymer, the single-cluster state becomes destabilized at $\tilde{\chi} < \tilde{\chi}^*$ and then smoothly disintegrates with heating; the behavior similar to that seen in Figure 2a.

c. Visual Analysis. The most direct approach to investigate the spatial organization of the microstructures observed in the simulation is to look at their snapshots. For the case $N = 127$, typical snapshots are shown in Figure 4a−d. At $\tilde{\chi} \lesssim \tilde{\chi}^*$ we see that the copolymer with alternating HA pattern forms two approximately equal clusters connected by a short stretched chain section (Figure 4a). Generally, we observed the same conformations for multiblock copolymers in the pretransition region, but for longer chains. With worsening solvent quality, both terminal clusters move toward the common center of mass by "eating" the stretched section. This leads to the formation of an elongated compact object (Figure 4b). It is the transition between the conformations of Figures 4a and 4b that corresponds to jumps in Figures 2 and 3.

On the other hand, proteinlike copolymer of a given length forms at $\tilde{\chi} \lesssim \tilde{\chi}^*$ an intermediate conformation with bulky head and long tail (Figure 4c); at larger values of N, two tails or long loop sections can be observed. When the solvent is made poorer, the globular core "swallows" the tail section, and the resulting microstructure has nearly spherical geometry (Figure 4d). The transition between the conformations shown in Figure 4c,d is also occurring with pronounced cooperativity, which is the reason for the inflection point in Figure 2a at high values of $\tilde{\chi}$.

It should be emphasized that the morphology of the final compact globule dramatically depends on copolymer sequence, especially for large N. As an example, we present in Figure 4e a snapshot picture illustrating the compact conformation of alternating copolymer with 255-unit hydrophobic chain at $\tilde{\chi} = 5.9$. It is seen that at very poor solvent quality for hydrophobic groups the polymer forms a cylindrical-shaped object. An analogous conformation is observed for long multiblock copolymers. Upon further visual inspection of these conformations, we find that the hydrophobic chain is in an irregularly folded (crumpled) state, forming the heart of the cylinder.

For intermediate values of $\tilde{\chi}$, the contraction of long alternating copolymers causes the formation of necklacelike structures, which are similar to that shown in Figure 4a, but these structures consist of the larger number of "pearls" of hydrophobic groups surrounded by hydrophilic groups. Generally, similar microstructures have been predicted under poor solvent conditions

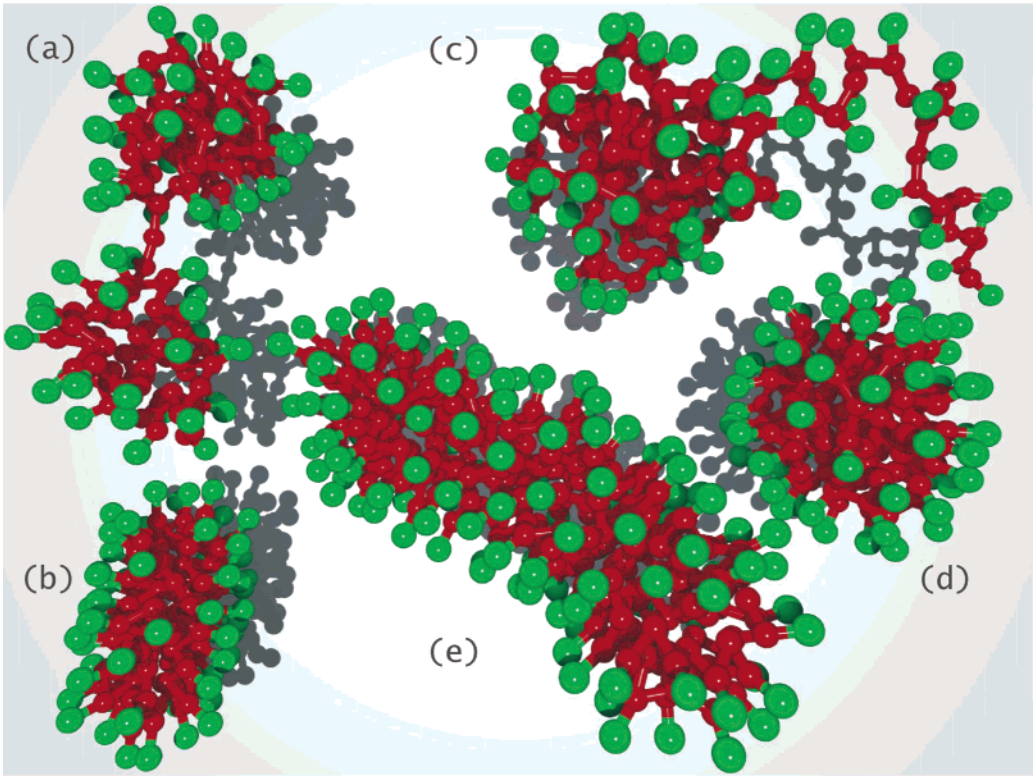

Figure 4. Snapshot pictures illustrating typical conformations of the amphiphilic copolymers: (a, b) alternating copolymer with hydrophobic chain length $N = 127$ at (a) $\tilde{\chi} = 2.71$ and (b) $\tilde{\chi} = 5.90$; (c, d) proteinlike copolymer with $N = 127$ at (c) $\tilde{\chi} = 1.10$ and (d) $\tilde{\chi} = 5.90$; (e) alternating copolymer with $N = 255$ at $\tilde{\chi} = 5.90$. Hydrophobic beads are shown as small red spheres, and hydrophilic beads are presented as larger green spheres. The sizes of all the spheres are schematic rather than space filling.

for amphiphilic homopolymers (poly-A) in our previous simulation.[21] They were observed in the present study also for sufficiently long multiblock copolymers. All these nonspherical microstructures, whose state is liquidlike, are thermodynamically stable.

For copolymers with proteinlike sequence we always observe the formation of dense spherical globules with well-separated hydrophobic and hydrophilic domains, similar to that shown in Figure 4d. Such behavior can be explained as follows. If we assume that the main driving force to form dense globules is a reduction of the exposed hydrophobic area, then one would expect spherically shaped aggregates to be energetically more favorable as they have the smallest area compared to volume. In fact, the sequence of proteinlike copolymers is specially tuned to form a spherical aggregate with a dense hydrophobic core and a hydrophilic shell. Nearly spherical aggregates are also possible for short regular multiblock copolymers.

On the other hand, large spherical aggregates from regular long-chain amphiphilic copolymers of the same HA composition are geometrically impossible without "structural defects", i.e., the inclusions of hydrophilic groups inside the hydrophobic core. In other words, small blocks along the chain cannot avoid unfavorable internal H−P contacts. In this case, the minimum of free energy can be attained only for nonspherical core geometries, when each hydrophobic group is close to core surface while all the hydrophilic side groups are positioned outside the hydrophobic core.[38,39] In other words, in an aggregate of elongated geometry, the energetically unfavorable inclusions of hydrophilic groups can be avoided, albeit at the cost of an increase in the surface free energy (for more details, see refs 21, 38, and 39). We will see below that these features are very important for understanding the aggregation processes in a multichain system.

d. Polymer Shape. To get a better feeling about the shape of the whole macromolecule, we also computed the gyration tensor describing the mass distribution given by

$$R_{ij} = (\delta_{ij}\delta_{kl} - \delta_{ik}\delta_{jl})\sum_{\alpha=1}^{\mathcal{N}}(x_{\alpha i} - x_{0i})(x_{\alpha j} - x_{0j}); \quad i, j, k, l = 1, 2, 3;\ i \neq l,\ j \neq k \quad (5)$$

where δ is the Kronecker delta, $x_{\alpha i}$ ($i = 1, 2, 3$) are the components of the position vector $\mathbf{r}_\alpha$ of the αth chain bead, and x_{0i} ($i = 1, 2, 3$) are the components of the center-of-mass (COM) vector $\mathbf{r}_0 = \mathcal{N}^{-1}\sum_\alpha \mathbf{r}_\alpha$, $\mathcal{N}$ being the total number of beads in the chain. The three eigenvalues of $\mathbf{R}$ are denoted by R_1^2, R_2^2, and R_3^2 (in ascending order); R_g^2 is the sum of these three eigenvalues. The ratios of these three eigenvalues, $f_1 = R_1^2/R_g^2$, $f_2 = R_2^2/R_g^2$, $f_3 = R_3^2/R_g^2$ ($f_1 \leq f_2 \leq f_3$), $k_1 = f_1/f_3$, and $k_2 = f_1/f_2$, determine the shape of the macromolecule. For a perfect sphere, the three eigenvalues are equal and $k_1 = k_2 = 1$. For an infinite cylinder, two of them are equal while the third is infinite, i.e., $k_1 \to 0$ and $k_2 = 1$. For a sheet (disklike object), one is finite and two are infinite so that $k_1 = k_2 \to 0$.

Figure 5 presents the ratios k_1 and k_2 as a function of $\tilde{\chi}$ for alternating and proteinlike copolymers with hydrophobic chain length $N = 31$ and 255. It is seen that in a thermodynamically good solvent ($\tilde{\chi} \lesssim 1$) the average shape of the swollen chains weakly depends on $\tilde{\chi}$ and corresponds to a three-axis ellipsoid, a result well-known for coillike chains.[40] An increase in $\tilde{\chi}$ is accompanied by well-pronounced variations in the polymer shape. For short chains, $N = 31$, as the solvent quality gets worse and the chain size decreases, we observe a structural transition from the initial coillike state to nearly spherical conformations with $k_1 \approx k_2 \to 1$, regardless of sequence. The proteinlike copolymer with relatively long backbone ($N = 255$) undergoes a smooth transition from elongated ellipsoid to a spherical object, while the alternating copolymer of the same length transforms into a stretched cylindrical-shaped conformation ($k_1 \to 1$ and $k_2 \ll 1$), which becomes more asymmetric as the effective solvent quality worsens. Of course, proteinlike globules are not perfect spheres due to thermal fluctuations, but their shape does not deviate strongly from a spherical form. By definition, the proteinlike copolymer itself is a prefabricated spherical aggregate. It should be also kept in mind that the shape of intermediate necklacelike structures predicted for regular block copolymers is quite close to cylindrical. This picture is clearly captured also in snapshots from simulations.

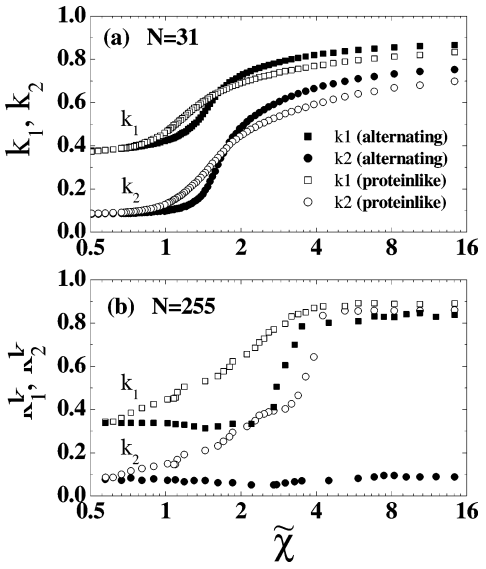

Figure 5. Ratios k_1 and k_2 as a function of the quality of the solvent, $\tilde{\chi}$, in a semilogarithmic scale for (■) alternating and (●) proteinlike copolymers with hydrophobic chain length (a) $N = 31$ and (b) $N = 255$.

Using the shape factors f_i, one can define the asphericity parameter, α_s, which is another useful quantification of the average shape

$$\alpha_s = \frac{1}{2}\sum_{i>j=1}^{3}(f_i - f_j)^2 \quad (6)$$

This value has a minimum, $\alpha_s = 0$, which occurs for a perfect sphere. It is nonzero for an ellipsoid, with a maximum, $\alpha_s = 1$, in the limit of a long thin cylinder. In cases where a cylinder begins to broaden, the asphericity parameter decreases. An aggregate can be considered spherical if $\alpha_s < 0.1$. Figure 6 presents the asphericity parameter for strongly collapsed alternating, multiblock and proteinlike copolymers with different hydrophobic chain length. We see that proteinlike copolymers in the collapsed state have a nearly spherical shape for all the chain lengths considered. The shape of regular copolymers begins to distinctly deviate from spherical at $N > 127$, and then it approaches a cylindrical shape with increasing chain length.

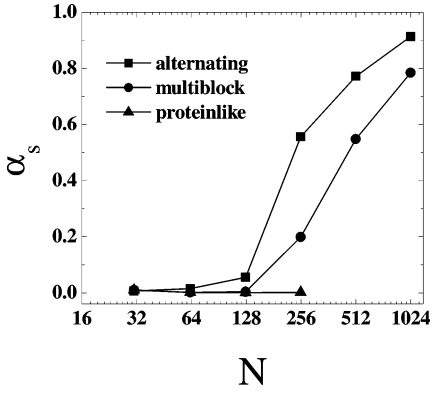

Figure 6. Asphericity parameter as a function of hydrophobic chain length, N, in a semilogarithmic scale for strongly collapsed (■) alternating, (●) multiblock, and (▲) proteinlike copolymers at $\tilde{\chi} = 23.0$.

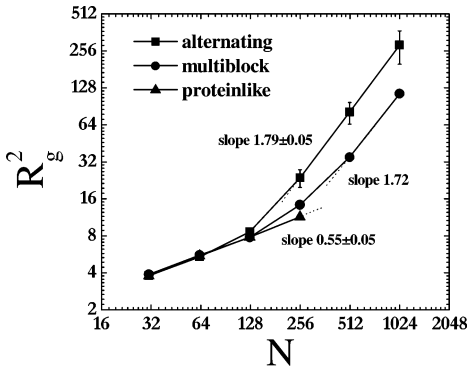

Figure 7. Mean-square radius of gyration R_g^2 as a function of the hydrophobic chain length N in a log−log scale for (■) alternating, (●) multiblock, and (▲) proteinlike copolymers at $\tilde{\chi} = 23.0$. The limiting slopes are shown near the corresponding curves.

e. *N* Dependence. For a homopolymer chain in the scaling regime (at $N \gg 1$), it is known that $R_g \propto N^{\nu}$, where ν depends on the solvent conditions: $\nu \approx 0.59$ for a swollen chain and $\nu = {}^1/_3$ for a collapsed chain.[1,2,41] However, taking into account the results discussed above, one can expect that the behavior of elongated cylindrical-shaped aggregates observed for regular amphiphilic copolymers in a poor solvent should be different. Indeed, as has been found in ref 21 for long-chain homopolymer amphiphiles, the formation of highly anisotropic conformations can lead to the $R_g \propto N^{0.9}$ scaling.

Figure 7 shows the mean-square radius of gyration R_g^2 as a function of the hydrophobic chain length N for the strongly collapsed alternating, multiblock and proteinlike copolymers at $\tilde{\chi} = 23.0$ in a log−log scale. For proteinlike copolymers with strong intrachain hydrophobic attraction, a fit to the simulation data yields a scaling exponent of $\nu \approx 0.28$, fairly close to the value of $\nu = {}^1/_3$ expected for strongly collapsed chains with a spherical geometry. On the other hand, for alternating and multiblock copolymers at $N > 255$, we find $\nu = 0.895$ and $\nu = 0.86$, respectively, somewhat less than the value corresponding to a rigid-rod-like object with $\nu = 1$.

It is clear that the cross section of a cylindrical-shaped aggregate is determined by the length of hydrophobic chain section separating amphiphilic sections. Because the dense cylindrical core exists as essentially a polymer melt, the hydrophobic sections of length L should obey the Flory theorem,[27] i.e., have a Gaussian size of the order of $L^{1/2}$ at $L \gg 1$. Therefore, the cylindrical cross section can be estimated as $\sigma L^{1/2}$, and the N-unit chain filling the hydrophobic core can be represented as an effective chain consisting of N/L renormalized monomer units of size $\sigma L^{1/2}$. Because of large stiffness of this renormalized "superchain", we expect that its size should behave as $R_g \propto \sigma N/L^{1/2}$ even at $N \gg 1$. The fact that the scaling exponents found here are smaller than this prediction is likely due to the fact that the values of N studied in our simulation are close to, though not yet in, this scaling regime. From the data of Figure 7, we see that the behavior observed for R_g tends to be closer to the $R_g \propto N$ scaling with increasing chain length.[42] It is easy to understand that for larger L the discussed asymptotic regime should be reached for larger N. That is why, almost for all the chain lengths considered in this paper for multiblock copolymers, we in fact deal with a crossover region.

2. Multichain System. In this section, we describe and compare the results of simulations for two multichain systems corresponding to alternating and proteinlike copolymers. Therefore, multiblock sequences are not considered here.

The multichain systems consisting of 127-unit alternating and proteinlike copolymers were simulated for the range of the $\tilde{\chi}$ parameter up to 4.06. As we have seen, under these conditions, single chains can form strongly collapsed conformations. For large values of $\tilde{\chi}$ distinctly larger than the critical value $\tilde{\chi}^*$ for globule formation, the multichain simulation can be trapped in metastable states. This is the reason why we have performed simulations only in the vicinity of the coil-to-globule transition temperature and not at temperatures far below this transition temperature.

To avoid metastable states, especially for large values of $\tilde{\chi}$ (or at low temperatures, $T \leq 1.0\ \epsilon/k_B$), most of the runs were started in different initial states. In some runs, the systems were started in a state with a random distribution of coillike chains inside the simulation box and $\tilde{\chi} = 0$. This system can be thought of as a reference system that is not "colored" yet, i.e., having the extra interaction (eq 2) between H and P sites set to zero, $\psi = 0$. Then, after a long equilibration, the interaction parameter $\tilde{\chi}$ was increased very slowly toward a desired level. In other runs the reverse procedure was employed. In this case, the system was started from an ordered configuration with collapsed chains and a strong interaction parameter $\tilde{\chi}$. A required thermodynamic state was then generated by decreasing $\tilde{\chi}$. Quantities relevant to the present study are the site−site pair correlation functions and static structure factors, providing information on density distribution and order in the system as well as the internal energy, the radius of gyration, etc.

a. Thermoreversibility. We found that for the simulated multichain systems the transitions between different temperature states corresponding to expanded and collapsed polymer conformations are thermodynamically reversible. As an example, let us consider the transitions from the low-temperature regime (the globular state, $T = 1.0\ \epsilon/k_B$) to the high-temperature regime (the coillike state, $T = 4.0\ \epsilon/k_B$) and then back to the low-temperature regime ($T = 1.0\ \epsilon/k_B$). After each of these runs, the temperature was monotonically raised (dropped) and again held constant for another (higher/lower) temperature for equilibration. Figure 8 shows the time evolutions of the reduced potential energy per particle, $v = U/\epsilon n\mathcal{N}$ (where U is the total potential

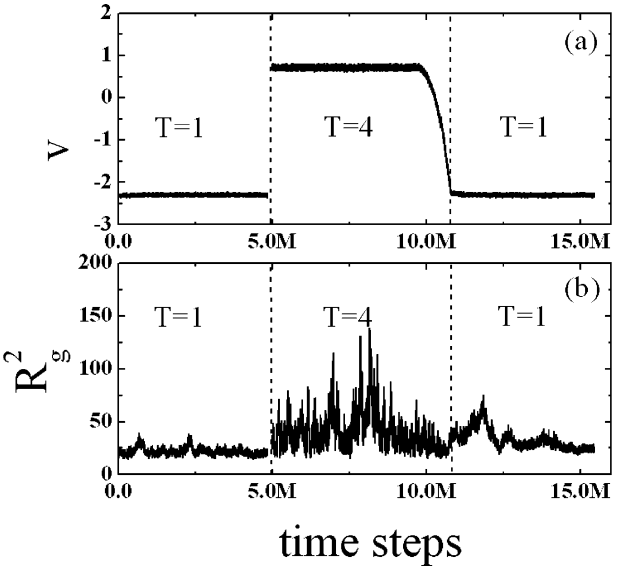

Figure 8. Time evolutions of (a) the reduced potential energy per particle, $v = U/\epsilon nN$ (where U is the total potential energy calculated at each time step) and (b) the squared radius of gyration for the multichain system composed of 127-unit amphiphilic copolymers with alternating distribution of H and A groups.

energy calculated at each time-step), and the squared radius of gyration for the multichain system composed of 127-unit amphiphilic copolymers with alternating distribution of H and A groups. As can be seen, the temperature transitions between the globular state and the coillike state are thermodynamically reversible, thereby indicating that the system under study does not get trapped in metastable states. The same is true for proteinlike copolymers. Below, we focus on the system morphology, which is characteristic of these temperature regimes.

b. Connolly Surfaces. For visual analysis of the simulated configurations, we employ the technique based on the construction of three-dimensional Connolly surfaces.[43,44] In this way, the global system morphology can be studied. The snapshots of the low-temperature configurations ($\tilde{\chi} = 4.06$) show that in this regime no large-scale aggregation of individual globules is observed for the multichain systems, implying that in this case the system lies in the stable one-phase region (Figure 9). This conclusion follows from very long MD simulations carried out at equilibrium. The same qualitative picture emerged when we repeated the simulation starting from different initial states: again, a stable array of globules was formed. In this respect, the behavior observed for the model amphiphilic copolymers is similar to that found for charged hydrophobic–hydrophilic proteinlike chains.[18] The simulation predicts the formation of specific microphase-separated morphologies in which strongly attracting hydrophobic chain sections form a distinct population of globules which are stabilized by a dense layer of hydrophilic beads. It is clear that the driving force for the microphase separation is competing interactions, that is, the strong attraction between the hydrophobic groups and repulsive interactions associated with the hydrophilic species. One may say that the intramolecular microphase separation prevents intermolecular aggregation, thus stabilizing the solution of globules.

Thus, under poor solvent conditions, we observe a stable solution of nonaggregating polymer globules which are well-separated from each other and form an array of colloidlike particles. Because of the fact that the amphiphilic globules are size- and shape-persistent objects, this allows them to maintain their morphological integrity even in concentrated enough solution.

c. Gyration Radii. From the partial mean-square radii of gyration calculated separately for hydrophobic and hydrophilic beads, R_{gH}^2 and R_{gP}^2, one can define the following characteristic ratio: $\zeta = R_{gP}^2/R_{gH}^2$ which takes into account both the properties of compactness and solubility for a heteropolymer globule.[5] (Compactness is directly related to the mean size of hydrophobic core, whereas solubility should dependent on the size of the hydrophilic shell preventing the aggregation.) In Figure 10, we show the ratio ζ as a function of the interaction parameter $\tilde{\chi}$. We find that this quantity is an increasing function of $\tilde{\chi}$. Qualitatively the same picture is observed for isolated chains. Such behavior is due to the fact that, as the attraction between H segments increases, the value of R_{gP}^2 decreases more slowly than R_{gH}^2, thus leading to demixing of H and P segments and facilitating their intramolecular microphase separation. This trend is more pronounced for

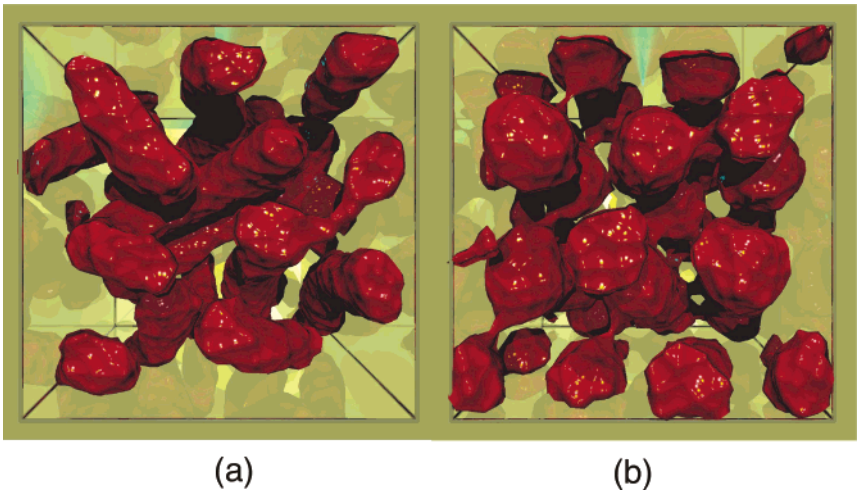

Figure 9. Snapshot pictures representing the three-dimensional Connolly surfaces generated for the low-temperature configurations ($\tilde{\chi} = 4.06$) of the multichain systems composed of the 127-unit amphiphilic copolymers with (a) alternating and (b) proteinlike distribution of H and A groups along the chain. The surfaces were generated only for the subsystem of hydrophobic chain beads, using a probe particle radius of $\sigma/2$. For visual clarity, hydrophilic side groups are intentionally not shown.

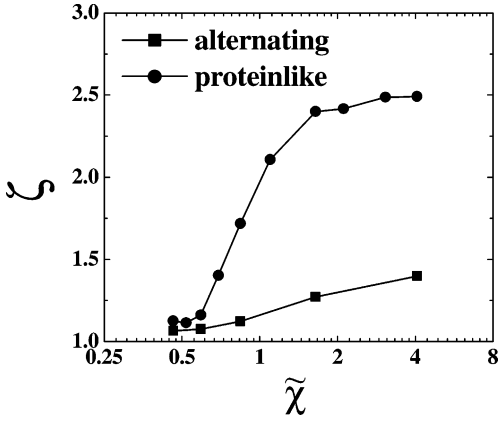

Figure 10. Ratio $\zeta = R_{gP}^2/R_{gH}^2$ as a function of the interaction parameter $\tilde{\chi}$ in a semilogarithmic scale for the 127-unit (■) alternating and (●) proteinlike chains in the corresponding multichain systems.

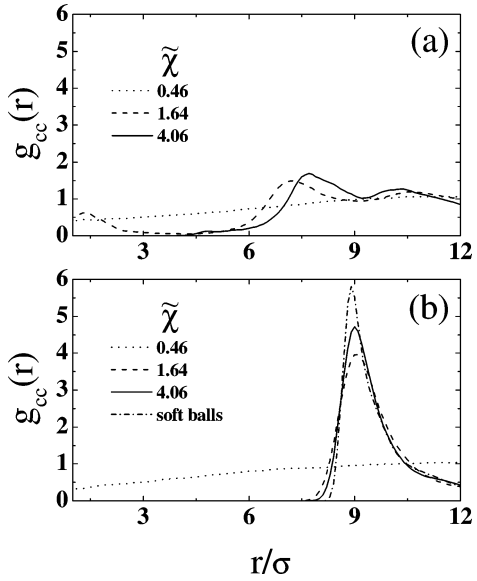

Figure 11. Center-of-mass pair correlation function $g_{cc}(r)$ for the multichain systems composed from (a) alternating and (b) proteinlike amphiphilic copolymers for three different values of the interaction parameter $\tilde{\chi}$. (b) The results obtained for the system of 27 "inflated balls" (see the text) are shown by the dash−dotted line.

the proteinlike copolymers than for the alternating copolymers, suggesting an idea that the former should be more protected against intermolecular aggregation.

d. Pair Correlation Functions. The local structure of the simulated systems was examined via different pair correlation functions $g(r)$ (PCF). They include the H−H, P−P, and chain−chain (CC) correlation functions. The latter describes the spatial distribution of centers-of-mass of copolymer chains. It helps us to understand the structural difference between two systems built up from cylindrical and spherical globules. Using the $g_{cc}(r)$ function, we can also estimate how close two globules may come to each other in a typical collision. Below, we will focus mainly on the features of this PCF.

Figure 11 shows the $g_{cc}(r)$ function for a few different values of $\tilde{\chi}$. Although the averaging was performed over very long runs, there is still substantial statistical noise in the simulation results, especially at short distances. For interpenetrating chains, $g_{cc}(r)$ is expected to be constant and equal to unity for all distances. However, we see that it exhibits nonrandom structure. In particular, when macromolecules are in coillike state ($\tilde{\chi} \leq 1.0$), there is a wide region at short distances where $g_{cc}(r)$ is smaller but comparable with unity. This means that the chains can interpenetrate, although not freely. Such behavior is well-known for interacting polymer coils in a good solvent (see, e.g., ref 45). At larger values of $\tilde{\chi}$ corresponding to globular state, we observe the development of a correlation hole with $g_{cc}(r) = 0$ for chain−chain distances r smaller than roughly 5σ and 7σ for the systems of alternating and proteinlike copolymers, respectively. The correlation hole becomes wider and moves to longer distances with increasing $\tilde{\chi}$. This fact indicates that polymers, being in globular state, cannot come closer together than this distance. Indeed, in this case the potential of mean force defined as $v_{cc}(r) = -k_B T \ln[g_{cc}(r)]$ becomes practically infinite. Thus, we can draw the following conclusion: under conditions corresponding to a poor solvent, when the chains are in a globular state, there is a very strong repulsive interaction pushing globules away from each other and thus preventing their merging into a single large cluster. This conclusion is supported by a visual inspection of a large number of snapshots similar to those shown in Figure 9.

Also, from Figure 11 it is seen that the behavior of the two systems under study is different. In particular, the $g_{cc}(r)$ function calculated for the system of alternating copolymers at $\tilde{\chi} > 1.6$ shows two peaks at short and larger interchain separations, while we find a single pronounced peak for the solution of proteinlike copolymers. Clearly, these features of the intermolecular pair correlation function are connected with the different three-dimensional packing of the collapsed chains having different conformation (spheres vs cylinders).

For the system of proteinlike copolymers, the following question arises: whether the packing of spherical globules leads to a liquidlike arrangement with significant structural fluctuation or an ordered body-centered-cubic (bcc) array of distinct spheres exists. Some aspects of this problem can be examined by comparing the simulation results with the predictions of simplified models. To this end, we considered the model of "inflated balls" that is a system of large soft spheres interacting via a shifted (repulsive) LJ potential (1) in which the parameter σ is set to a large value. We have simulated the system composed of $n = 27$ balls enclosed into a periodic box of the same size as in the calculations discussed above. During the initial phase of the simulation, the size of the balls was progressively increased from $\sigma = 1$ to $\sigma = 10$ so that after equilibration the system had nearly bcc structure. Note that in this simulation the strongly repulsive large balls can only fluctuate around its average positions but cannot diffuse. The results are shown in Figure 11b where we present the pair correlation function $g(r)$. One sees that the PCF obtained for this oversimplified model is close to the center-of-mass PCF found for the system of proteinlike copolymers at poor solvent quality. In particular, the positions of the main peaks, r_1, are practically coinciding, $r_1 = 8.9\sigma$. Therefore, the structural organization of the polymer system under these conditions is close to a slightly eroded bcc array formed by soft "inflated balls". Similar morphologies are characteristic of strongly charged particles (e.g., they have been observed for dense one-component plasma at very strong electrostatic repulsion[46]).

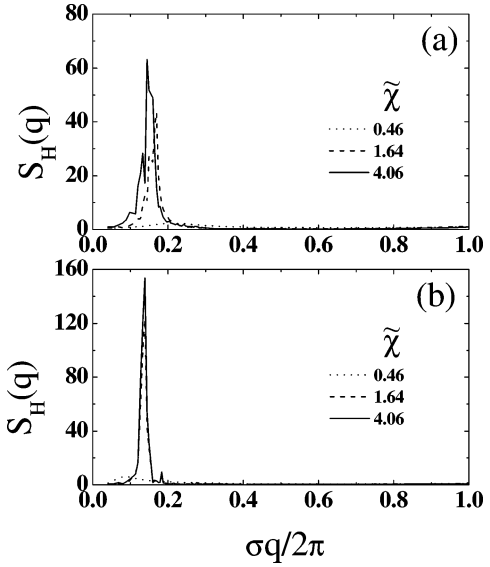

Figure 12. Partial static structure factors, $S_H(q)$, calculated for hydrophobic chains of the multichain systems composed from (a) alternating and (b) proteinlike amphiphilic copolymers for three different values of the interaction parameter $\tilde{\chi}$.

e. Structure Factors. Further evidence for the formation of the stable nonaggregated structure of globules similar to strongly interacting charge-stabilized suspensions or charged microgels[19,20] can be gained from the analysis of structure factors, $S(q)$, which characterize the spatial organization of the system as a whole. We have calculated the partial static structure factor, $S_H(q)$, that reflects the arrangement of the hydrophobic component in space and is defined as a decomposition of the total structure factor in the following way:

$$S_H(q) = \frac{1}{N_H}\sum_{\alpha=1}^{N_H}\sum_{\beta=1}^{N_H}\langle \exp(i\mathbf{q}\cdot\mathbf{r}_{\alpha\beta})\rangle \qquad (7)$$

where $i = \sqrt{-1}$, q is the wavenumber, and $N_H = nN$ is the number of hydrophilic sites in the system. The normalization for $S_H(q)$ is $S_H(0) = k_B T \rho_H \chi_T$, where $\rho_H = N_H/V$ is the partial number density and χ_T denotes the compressibility.

The function $S_H(q)$ is presented in Figure 12 for three different values of $\tilde{\chi}$. Although there are technical problems connected with the calculation of $S_H(q)$, due to the fact that it is strongly influenced by finite size effects (especially at small q), from Figure 12 we can definitely see that for both systems the function $S_H(q)$ has an intensive "small-angle" peak for the regime corresponding to a poor solvent. At $\tilde{\chi} < 1$, the magnitude of H−H correlations appears to be relatively insensitive to $\tilde{\chi}$ and negligibly small as compared to that observed for poor solvent conditions. The main maximum of $S_H(q)$ implies that the systems are spatially inhomogeneous at the intermediate scale related to the average interchain distance. The scattering intensity increases as the solvent becomes poorer. The position of the main peak, q^*, is determined by the interchain correlations at relatively long distances and connected with the correlation holes seen in Figure 11 for the chain−chain PCF. Indeed, the characteristic wavenumber q^* determines the average distance r^*, given by the Bragg condition $r^* = 2\pi/q^*$, between scattering objects; e.g., at $\tilde{\chi} = 4.06$ we find $r^* = 6.9\sigma$ and $r^* = 7.2\sigma$ for the alternating and proteinlike copolymer systems, respectively, the values which are close to the size of the corresponding correlation holes (see Figure 11). If we assume a uniform distribution of nonaggregated spherical globules in the simulation box, we obtain the following estimation: $r^* = L_{box}\sqrt[3]{n} = 8.33\sigma$. This value is close to the results presented above. On the other hand, for uniformly distributed dimers formed by the aggregated globules this distance would distinctly increase ($r^* = L_{box}/\sqrt[3]{n/2} \approx 11\sigma$) and would become even longer for trimers and larger aggregates. All these findings indicate that in this temperature regime we indeed deal with a system of single-chain nonaggregated (or weakly aggregated) globules, which repel each other because they are covered with a dense hydrophilic "skin", as it has been shown in the previous subsection.

Therefore, we observe no large-scale aggregation or phase separation for a system containing amphiphilic globules simulated with our side-chain HA model. This is in noticeable contrast with the predictions for amphiphilic linear HP models[6−9,47] which do not treat hydrophilic side groups explicitly.

f. Aggregation Numbers. Although the existence of separated single-chain globules is a prevalent structural motif for both systems in a poor solvent, nevertheless, we observed the episodical formation of intermolecular aggregates: when the temperature becomes progressively lower, the amphiphilic globules ("unimers") can stick together, forming pairs, trimers, and (very seldom) multimers.[48] Association and dissociation is an equilibrium process leading to a size distribution of the intermolecular aggregates. To describe this process, we analyze the average aggregation numbers.

In the simulation, an amphiphilic chain is considered to belong to an intermolecular aggregate if any of its hydrophobic part touches a neighboring hydrophobic part from another chain. If there are no such intermolecular contacts present, the chain is considered as a unimer. The size of an aggregate defined in this way is described by the aggregation number $\langle M\rangle$, which is the average number of macromolecules in the aggregate. This quantity can be calculated as $\langle M\rangle = \sum_M MW(M)$ or as $\langle M\rangle = \langle m\rangle/N$, where $W(M)$ is the cluster size distribution function, $\langle m\rangle$ is the total average size of the hydrophobic cluster present in the simulation box, n is the number of the chains, and N is the length of hydrophobic chain. Obviously, without any attraction between hydrophobic sites (i.e., at $\tilde{\chi} = 0$) the time-averaged cluster size distribution function and the corresponding mean aggregation number are determined by random contacts between these sites and depend primarily on their number density. In this case, one can expect that $\langle M\rangle \ll 1$. For an infinitely dilute solution of nonaggregated globules (at $\tilde{\chi} \gg 1$) we anticipate that $\langle M\rangle \approx 1$. The formation of intermolecular aggregates would result in $\langle M\rangle > 1$.

Taking into account the estimates given above for $\langle M\rangle$, let us look at the simulation data presented in Figure 13 for the two systems simulated at different values of $\tilde{\chi}$. We see that with the "strict" definition of hydrophobic cluster used in our calculation ($r \le r_0$, cf. eq 1) the proteinlike globules practically show no intermolecular aggregation: for the whole range of values of $\tilde{\chi}$ considered, we find that $\langle M\rangle \le 1$. Of course, the value of $\langle M\rangle$ is strongly affected by the particular choice for the contact distance r_0 (it increases with r_0). In contrast to the proteinlike chains, we observe that the alternating copolymers tend to form small aggregates with $\langle M\rangle \ge 2$

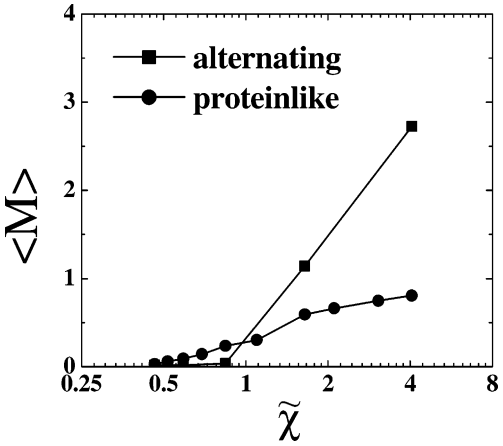

Figure 13. Average aggregation number as a function of the quality of the solvent, $\tilde{\chi}$, in a semilogarithmic scale for the multichain systems of (■) alternating and (●) proteinlike copolymers.

when the hydrophobic attraction becomes sufficiently strong. Thorough inspection of the snapshot pictures corroborates this observation. From the snapshots it is also seen that the cylindrical globules can stick together by their ends, forming elongated multiglobular structures composed from stable individual globules connected between each other in a linear fashion.

g. Aggregation Free Energy. An alternative way to study the process of cluster formation is to monitor the change in the corresponding free energy, ΔG. To this end, following the standard quasi-chemical approach,[49] we can treat the polymer solution as a multicomponent system, where intermolecular aggregates of different size (A_M, $M > 1$) are present in equilibrium with unimers (A_1), $A_1 \rightleftarrows A_M$. These species are treated as distinct chemical components, each characterized by its own solution concentration and chemical potential. The concentrations $[A_1]$ and $[A_M]$ of the species (or their mole fractions) can be found via the integration of the center-of-mass PCF; this gives an estimate for the overall association equilibrium constant K and ΔG, which is the difference in the standard Gibbs free energy between chains belonging to intermolecular aggregates and unimers (for more computational details, see the Appendix). When $\Delta G < 0$, it is energetically favorable for the chains to merge. If $\Delta G > 0$, the unimers are favorable.

The calculated values of ΔG are presented in Figure 14 as a function of the interaction parameter $\tilde{\chi}$. As seen, lowering the temperature, or equivalently, increasing the interaction parameter $\tilde{\chi}$, shifts the equilibrium $A_1 \rightleftarrows A_M$ ($M \geq 2$) in the system of proteinlike copolymers toward the nonaggregated state, and this is reflected in Figure 14 as an increase in ΔG. At first sight, such behavior is somewhat counterintuitive. Actually, from a general consideration it would be possible to expect an opposite behavior, when worsening in the solvent quality facilitates the aggregation and thus reduces the aggregation free energy. However, the observed behavior becomes quite clear if one takes into account the results discussed above. An increase in the free energy is explained by the appearance of dense hydrophilic shell around the formed spherical globules, which serves as a practically insurmountable energy barrier preventing the aggregation. When the solvent becomes poorer, chains are compressed, and this is accompanied by strengthening this hydrophilic "protective barrier". On

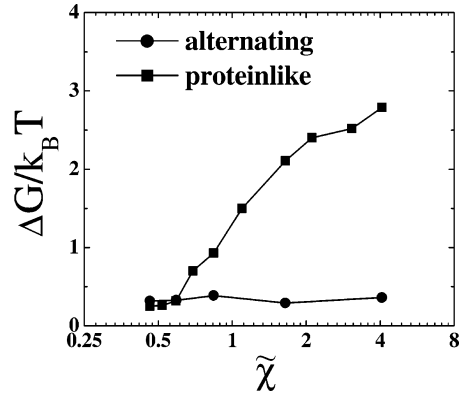

Figure 14. Aggregation free energy ΔG as a function of the interaction parameter $\tilde{\chi}$ in a semilogarithmic scale for the multichain systems of (■) alternating and (●) proteinlike copolymers.

the other hand, the free energy of aggregation estimated for the system of alternating copolymers weakly depends on the solvent quality and remains close to zero for all values of the energy parameter $\tilde{\chi}$ examined here (see Figure 14). This behavior also can be understood on the basis of the data about the conformational structure of this copolymer. Since at the same HA composition the cylindrical globules have the larger surface-to-volume ratio, the hydrophilic shell is not so dense, and therefore, it does not ensure a sufficient protection, while for the proteinlike copolymers rather high densities of the hydrophilic shell are reached. Such a behavior can be also expected from the data presented in Figure 10. Moreover, the thorough analysis of globular conformations shows that this layer is almost absent near the faces of the cylinder. These facts explain the more expressed tendency of regular copolymers toward the aggregation in solution. Nevertheless, we emphasize that, similar to proteinlike copolymers, in this case the large-scale aggregation of macromolecules is not observed, at least for the range of the parameters and the time scales studied here.

IV. Concluding Remarks

In this paper we have developed a minimal ("black-and-white" or, more precisely, "black-and-gray") model of hydrophobic/amphiphilic (HA) copolymers, with the understanding that many features not present in it can later be easily incorporated. In contrast to the standard two-letter "beads-on-a-string" HP model,[3,4] our "side-chain" HA model incorporates the dualistic nature of amphiphilic monomer units, each consisting of hydrophobic and hydrophilic interaction sites. With this model, we have reported on the results of molecular dynamics simulations of the hydrophobically driven conformational transitions under the variation of solvent conditions, both for single-chain and for multichain systems. A principal conclusion from the model is that many of the general features of amphiphilic copolymers can be predicted from simple and nonspecific interaction potentials.

We have simulated single HA copolymers with the same HA composition but with different distributions of H and A monomer units along the main hydrophobic chain, viz., regular copolymers comprising H and A units in alternating sequence, regular multiblock copolymers composed of H and A blocks of equal lengths, and the quasi-random proteinlike copolymers[13] having quenched primary structure. Under poor solvent conditions for

hydrophobic segments, all the copolymers form compact conformations, irrespective of the primary structure. However, the morphology of these conformations dramatically depends on copolymer sequence, especially for long chains. We have shown that single proteinlike polyamphiphiles can readily adopt conformations of compact spherical globules with the hydrophobic chain sections clustered at the globular core and the hydrophilic side groups forming the envelope of this core and buffering it from polar solvent. This morphology closely resembles that of micelles[50] or globular proteins. For all the chain lengths studied, these structures are nearly spherical with small fluctuations. For the range of the hydrophobic chain lengths N simulated in this study ($N \leq 255$), we have demonstrated that the chain size R_g as a function of N behaves as $R_g \propto N^{\nu}$ with $\nu \approx 0.28$, the exponent expected for collapsed chains.[1,2,41]

The globules of relatively short regular multiblock copolymers (with a fixed block length of $L = 3$ and $N \lesssim 127$) are also spherical or nearly so. On the other hand, at poor solvent conditions, the compact conformations of long regular copolymers tend to be elongated in one direction, especially for alternating HA sequence. The hydrophobic core formed by these copolymers increases with chain length, in a manner that can be understood on the basis of a uniform core which expands with chain length in one direction, and a sharp hydrophobic/hydrophilic interface whose width is essentially constant. We have found that for sufficiently long chains the formation of such conformations can lead to the $R_g \propto N^{\nu}$ scaling with $0.86 \leq \nu \leq 0.89$, although the range of N that we could study was too small to draw definite quantitative conclusions about this dependence. In our previous simulations we have predicted the $R_g \propto N^{0.9}$ scaling for homopolymer amphiphiles with $N \leq 1024$.[21] This behavior is connected with the formation of strongly elongated core−shell conformations having a locally cylindrical symmetry. This is due to strong intramolecular segregation that reduces the number of energetically unfavorable H−P contacts. Certainly, the steric (excluded volume) interactions between hydrophilic side groups also play an important role. In this connection we would like to note that the side chains (groups) incorporated in the minimalist linear models of proteins can introduce more topological frustration but also serve to steer model proteins into the proper configuration and thus enhance the cooperativity of folding.[51−53]

We may add a few words about experimental evidences of some facts predicted in the present simulation. Kikuchi and Nose[54] have reported on a systematic experimental studies of poly(methyl methacrylate)-*graft*-polystyrene (PMMA-*g*-PS) with short branches in a selective solvent (isoamyl acetate) which is a good solvent for PS. Under given solvent conditions, this copolymer behaves as an amphiphilic copolymer, bearing a resemblance to the model considered in our simulation. At high branch density, the authors[54] have observed the formation of thermodynamically stable unimolecular rodlike micelles formed via intramolecular segregation of the PMMA backbone and PS branches, with the shrunken PMMA backbone making the rodlike core covered with PS chains. Also, it has been found that the rod is not necessarily rigid, but may be flexible in the weakly segregated state, and becomes more rigid with stronger segregation upon decreasing temperature, i.e., upon the progressive worsening of solvent quality for the PMMA backbone. Selb and Gallot[55] have demonstrated that poly(styrene)-*graft*-poly(4-vinyl-*N*-ethylpyridium bromide) forms unimolecular micelles in water/methanol mixtures. These experimental results can be treated as an indirect confirmation of our conclusion that sufficiently long regular copolymers with amphiphilic monomer units do form intramolecular anisometric micellar structures in a poor solvent. The presence of stable single-chain core−shell nanostructures in solution of amphiphilic copolymers has also been observed by Wu and Qiu.[56] Using a combination of static and dynamic laser light scattering, they have found that a linear poly(*N*-isopropylacrylamide) chain grafted with poly(ethylene oxide) (PNIPAM-*g*-PEO) in water can undergo a coil-to-globule transition to form spherical single-chain aggregates with collapsed PNIPAM chain backbone as the hydrophobic core and the grafted short PEO chains as the hydrophilic shell. In general, these colloidlike nanostructures are similar to those observed in our simulation for proteinlike amphiphilic copolymers.

We would like to emphasize that the physical picture observed in the present simulation for coil-to-globule transition, at least for regular copolymers, is generally very close to that known for processes of micellization in dilute solutions of surfactants.[50,57] In fact, our HA model displays a behavior that is rather common for usual polysoaps that, due to their amphiphilic character, are able to build up a variety of self-assembled structures in solution.[58−60] In particular, the intermediate necklacelike structure predicted for long regular amphiphilic copolymers may be considered as a string of intramolecular micelles.[61] Because of strong repulsive interaction between their outer hydrophilic regions, there is an energy barrier associated in bringing two separated micelles together. It is clear that such a situation is very similar to that characteristic of usual low-molecular-weight surfactants, which form micelles in a dilute solution.[50] Unlike ordinary micelles, however, intramolecular micelles need no critical concentration of polysurfactants for their formation because under the connectivity constraints there is no loss of translational entropy. In other words, the configurational freedom of a collection of amphiphilic groups is much more limited when they are linked monomer units along a chain than when they are detached independent surfactants. One can say that the connectivity of the chain, on which the amphiphilic groups are attached, forces the concentration of these groups to be locally very high even at bulk polymer concentrations tending to zero. Chain connectivity also imposes constraints on the packing of the amphiphilic groups in compact conformations.

Also, we simulated the solution of short-chain alternating and proteinlike copolymers. Because of computational difficulties, we have examined solutions only with one concentration but at different temperatures (solvent qualities). We intentionally assumed that the solution concentration is sufficiently high so that the chains in the globular state had a nonzero probability of coming in contact. A study of the aggregation tendencies of copolymers in a poor solvent was the main purpose of this simulation. The transitions between different temperature regimes have been found to be thermodynamically reversible. This means that the structural morphologies, corresponding to coillike and globular conformations observed in the system, are reversibly restored with a sequential change in the temperature.

Our main conclusion is that amphiphilic proteinlike macromolecules do not manifest a visible tendency toward the aggregation for the investigated conditions when single chains form compact globules. In other words, we have observed the thermodynamically stable solution of nonaggregated globules even at relatively high polymer concentration. Generally speaking, the reason for this behavior is simple. It is known that low-molecular-weight surfactants dramatically increase the stability of polymers and are widely used to prevent aggregation in polymer solutions.[50] In our model, "surfactants", i.e., amphiphilic A groups, are incorporated into polymer chain, thus ensuring the stabilizing effect. From the temperature dependencies of chain size and aggregation free energy we can conclude that below the collapse transition temperature there is free energy barrier preventing the aggregation of copolymer globules. Therefore, in a macroscopic system, precipitation of a macroscopic polymer-rich phase should be suppressed. This is quite different from the solution of usual linear polymers (both homo- and heteropolymers), where the aggregation process in a poor solvent is not associated with a free energy cost and the aggregation is taking place together with the single-chain collapse transition.[41] The situation with the alternating copolymers is not so clear. Although in this case we also did not observe the large-scale aggregation, which in a macroscopic system has to be accompanied by polymer precipitation, such copolymers were found to be capable of forming sufficiently large intermolecular aggregates. We explain these differences by the fact that the spherical and cylindrical globules, formed by these two types of copolymers, have hydrophilic shells with noticeably different density. Obviously, aggregation should be less likely in a more dilute solution.

As has already been mentioned in the Introduction, in recent years different groups tried to design nonaggregating heteropolymers.[5−12] However, until now such studies did not give a single valued positive result. In our opinion, the basic reason for such failures lies in the fact that in these works a study was limited only to ordinary linear heteropolymer chains. Meanwhile, it is far from clear that the desired result can be in principle accessible on the basis of simple linear models in the case of electroneutral heteropolymers, since in this case there are no real ways of creating the insurmountable (or very high) energy barriers, preventing large-scale aggregation. It is interesting that this conclusion is in general agreement with the recent study[62] where it was shown (both experimentally and theoretically) that branched polymer amphiphiles are normally much better surfactants than traditional linear analogues.

There are many further issues that can be addressed by a computer technique and the model of the kind presented here. It will be worthwhile to examine the behavior of significantly longer amphiphilic chains as well as the chains with rigid (semirigid) hydrophobic backbone. The corresponding calculations are currently being performed. An interesting issue is phase behavior and transitions between different morphologies in dilute and semidilute solutions. Of course, to compute phase diagrams, a grand canonical ensemble is better than the canonical ensemble used in this simulation, although this step would entail vastly more computational effort. One can also introduce dipole moments for side H−P bonds and specific interactions (like hydrogen bonds) for some of the chain units. It is expected that these additional factors can result in the formation of intramolecular secondary structures and lead to further increase in globule stability.

Appendix

The aggregation of amphiphilic molecules into intermolecular clusters can be described, to first order, using the general framework of reversible coagulation process.[63] We consider a volume V, which contains n amphiphilic molecules. At any time, some fraction of the n copolymer chains can be self-assembled into interchain aggregates of different sizes so that in the system there are n_1 free chains with concentration $[A_1]$, n_2 dimers with concentration $[A_2]$, and so on. They must satisfy the condition $\sum i n_i = n$. The aggregation in a solution can be represented by the multiple-equilibrium model[49,64−66] described by

$$A_1 \overset{+A_1}{\rightleftharpoons} A_2 \overset{+A_1}{\rightleftharpoons} \ldots \overset{+A_1}{\rightleftharpoons} A_M \overset{+A_1}{\rightleftharpoons} \ldots \qquad \text{(A1)}$$

with the following kinetic equations

$$\partial[A_1]/\partial t = -k_1[A_1]^2 + k_2[A_2]$$

$$\partial[A_2]/\partial t = -k_1[A_2][A_1] + k_2[A_3] + k_1[A_1]^2 - k_2[A_2]$$

$$\ldots$$

$$\partial[A_M]/\partial t = -k_1[A_M][A_1] + k_2[A_{M+1}] + k_1[A_{M-1}][A_1] - k_2[A_M] \qquad \text{(A2)}$$

where $[A_1]$ is the time-dependent concentration of single chains (unimers), $[A_i]$ is the time-dependent concentration of aggregates containing i chains, and k_1 and k_2 are the rate constants of aggregation and dissociation, respectively. At equilibrium, we have

$$-k_1[A_1]^2 + k_2[A_2] = 0$$

$$-k_1[A_2][A_1] + k_2[A_3] + k_1[A_1]^2 - k_2[A_2] = 0$$

$$\ldots$$

$$-k_1[A_M][A_1] + k_2[A_{M+1}] + k_1[A_{M-1}][A_1] - k_2[A_M] = 0 \qquad \text{(A3)}$$

Here and below, $[A_i]$ denote the corresponding equilibrium concentrations of species present in the system. Using the definition for the overall association equilibrium constant, $K = k_1/k_2$, from eq A3 one can obtain

$$K = \frac{1 - \sqrt{1-\phi}}{[A_1]} = \frac{1 - \sqrt{1-\phi}}{C(1-\phi)} \qquad \text{(A4)}$$

where $\phi = C_A/C$ is the fraction of the chains entering aggregates, C_A is the concentration of the chains in all the aggregates that can be found in the system, and C is the total chain concentration. Finally, for the aggregation free energy, we have

$$\Delta G/k_B T = -\ln\left(\frac{1 - \sqrt{1-\phi}}{C(1-\phi)}\right) \qquad \text{(A5)}$$

This free energy characterizes the aggregates as a "pseudophase" in thermal equilibrium with unimers.

The fraction of the chains entering intermolecular aggregates, ϕ, can be estimated from the center-of-mass

pair correlation function, $g_{cc}(r)$. Having $g_{cc}(r)$, the corresponding running integration numbers, $n(r)$, can be found. They are defined as

$$n(r) = 4\pi\rho_c \int_0^r g_{cc}(x)x^2\,dx \qquad \text{(A6)}$$

where ρ_c is the total number density of the chains, n/V. The value of $n(r)$ is the average number of the chains around a given chain at a distance r. We assume that two chains belong to an interchain aggregate if their center-of-masses are located close to each other than some characteristic aggregation distance R_a. In other words, the $n(R_a)$ chains form a connected cluster. The volume fraction of intermolecular aggregates larger than unimers, $\phi_{(M>1)} = \sum_{i=2}^{M}\phi_i$, can be calculated from the relation

$$\phi = \int_0^{R_a} g_{cc}(r)r^2\,dr \Big/ \int_0^{L_{box}/2} g_{cc}(r)r^2\,dr \qquad \text{(A7)}$$

where $R_a \le L_{box}/2$ and L_{box} is the size of the simulation box with periodic boundary conditions. Certainly, the choice of R_a for a dense multichain system is rather arbitrary. For simplicity, we have chosen the parameter R_a to be equal to the characteristic distance between noninteracting particles uniformly distributed at the number density of n/L_{box}^3 that, in our case, gives $R_a = L_{box}/\sqrt[3]{n} = 8.33\sigma$. This "mean-field-like" state, which corresponds to $g_{cc}(r) = 1$, is considered as a reference state.

Acknowledgment. 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 (Project 04-03-32185) is highly appreciated.

References and Notes

(1) Lifshitz, I. M.; Grosberg, A. Y.; Khokhlov, A. R. *Rev. Mod. Phys.* **1978**, *50*, 683.
(2) Grosberg, A. Y.; Khokhlov, A. R. *Statistical Physics of Macromolecules*; American Institute of Physics: New York, 1994.
(3) Lau, K. F.; Dill, K. A. *Macromolecules* **1989**, *22*, 3986.
(4) Lau, K. F.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6388.
(5) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 839.
(6) Broglia, R. A.; Tiana, G.; Pasquali, S.; Roman, H. E.; Vigezzi, E. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12930.
(7) Giugliarelli, G.; Micheletti, C.; Banavar, J. R.; Maritan, A. *J. Chem. Phys.* **2000**, *113*, 5072.
(8) Timoshenko, E. G.; Kuznetsov, Y. A. *J. Chem. Phys.* **2000**, *112*, 8163.
(9) Bratko, D.; Blanch, H. W. *J. Chem. Phys.* **2001**, *114*, 561.
(10) Bratko, D.; Blanch, H. W. *J. Chem. Phys.* **2003**, *118*, 5185.
(11) Gupta, P.; Hall, C. K.; Voegler, A. C. *Protein Sci.* **1998**, *7*, 2642.
(12) Toma, L.; Toma, S. *Biomacromolecules* **2000**, *1*, 232.
(13) Khokhlov, A. R.; Khalatur, P. G. *Physica A* **1998**, *249*, 253.
(14) Khokhlov, A. R.; Khalatur, P. G. *Phys. Rev. Lett.* **1999**, *82*, 3456.
(15) Virtanen, J.; Baron, C.; Tenhu, H. *Macromolecules* **2000**, *33*, 336.
(16) Virtanen, J.; Tenhu, H. *Macromolecules* **2000**, *33*, 5970.
(17) Virtanen, J. In *Self-assembling of Thermally Responsive Block and Graft Copolymers in Aqueous Solutions (Ph.D. Thesis)*; Department of Chemistry, University of Helsinki: Helsinki, 2002; p 42.
(18) Khalatur, P. G.; Khokhlov, A. R.; Mologin, D. A.; Reineker, P. *J. Chem. Phys.* **2003**, *119*, 1232.
(19) Micka, U.; Holm, C.; Kremer, K. *Langmuir* **1996**, *12*, 4033.
(20) Micka, U.; Kremer, K. *Europhys. Lett.* **2000**, *49*, 189.
(21) Vasilevskaya, V. V.; Khalatur, P. G.; Khokhlov, A. R. *Macromolecules* **2003**, *36*, 10103.
(22) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1990.
(23) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24.
(24) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237.
(25) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227.
(26) Khalatur, P. G.; Novikov, V. V.; Khokhlov, A. R. *Phys. Rev. E* **2003**, *67*, 051901.
(27) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953.
(28) While the different relative strengths of H−H, H−P, and P−P interactions determine the segregation of H and P beads in compact (globular) conformations, the integral interaction parameter (eq 4) controls the degree of compactness of these conformations and also the aggregation properties of copolymers.
(29) Khalatur, P. G.; Balabaev, N. K.; Pavlov, A. S. *Mol. Phys.* **1986**, *59*, 753.
(30) Sullivan, F.; Mountain, R. D.; O'Connell, J. *J. Comput. Phys.* **1985**, *61*, 138.
(31) Yamakawa, H. *Modern Theory of Polymer Solutions*; Harper and Row: New York, 1971.
(32) Govorun, E. N.; Ivanov, V. A.; Khokhlov, A. R.; Khalatur, P. G.; Borovinsky, A. L.; Grosberg, A. Y. *Phys. Rev. E* **2001**, *64*, 040903.
(33) Van den Oever, J. M. P.; Leermakers, F. A. M.; Fleer, G. J.; Ivanov, V. A.; Shusharina, N. P.; Khokhlov, A. R.; Khalatur, P. G. *Phys. Rev. E* **2002**, *65*, 041708.
(34) Ptitsyn, O. B. In *Protein Folding*; Creighton, T. E., Ed.; W.H. Freeman: New York, 1992; p 253.
(35) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Rev. Mod. Phys.* **2000**, *72*, 259.
(36) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
(37) Baiesi, M.; Carlon, E.; Orlandini, E.; Stella, A. L. *Phys. Rev. E* **2001**, *63*, 041801.
(38) Khalatur, P. G.; Khokhlov, A. R.; Nyrkova, I. A.; Semenov, A. N. *Macromol. Theory Simul.* **1996**, *5*, 713.
(39) Khalatur, P. G.; Khokhlov, A. R.; Nyrkova, I. A.; Semenov, A. N. *Macromol. Theory Simul.* **1996**, *5*, 749.
(40) Solc, K.; Stockmayer, W. H. *J. Chem. Phys.* **1971**, *54*, 2756.
(41) De Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
(42) Although any extrapolation toward $N \to \infty$ clearly would be very subtle, it is believed that in the $N \to \infty$ limit, because of strong thermal fluctuations, the limiting scaling exponent would be close to that expected for the good solvent regime, due to repulsive interactions between outer hydrophilic groups shielding hydrophobic core in the resulting core−shell cylindrical structure. For the $N \to \infty$ limit, we expect to find $R_g \propto \sigma L^{1/2}(N/L)^{3/5} \propto \sigma N^{3/5}/L^{1/10}$. In this limit, the locally cylindrical "superchain" as a whole would look like a coiled nonintersecting hose having a finite thickness and a finite persistent length. Of course, this regime is far beyond any reasonable computer facilities.
(43) Connolly, M. L. *Science* **1983**, *221*, 709.
(44) A Connolly surface is the locus of points formed by the intersection of the van der Waals surface of a given ensemble of particles with a spherical probe particle of radius, which is rolled over the particles under consideration. The Connolly surfaces were generated for the subsystem of hydrophobic chain beads, using a probe particle radius of $\sigma/2$.
(45) Grosberg, A. Y.; Khalatur, P. G.; Khokhlov, A. R. *Makromol. Chem. Rapid Commun.* **1982**, *3*, 709.
(46) Ng, K.-C. *J. Chem. Phys.* **1974**, *61*, 2680.
(47) Zherenkova, L. V.; Talitskikh, S. K.; Khalatur, P. G.; Khokhlov, A. R. *Dokl. Phys. Chem.* **2002**, *382*, 23.
(48) In should be noted that the use of periodic boundary conditions means that it is possible to form periodic aggregates as a way of avoiding the creation of an additional interface, thus increasing the ratio of volume to surface area.
(49) Mukerjee, P. *J. Phys. Chem.* **1972**, *76*, 565−570.
(50) Mittal, K. L., Ed. *Micellization, Solubilization, and Microemulsions*; Plenum Press: New York, 1977; Vols. 1 and 2.
(51) Li, M. S.; Klimov, D. K.; Thirumalai, D. *Comput. Phys. Commun.* **2002**, *147*, 625.
(52) Kussell, E.; Shimada, J.; Shakhnovich, E. I. *J. Mol. Biol.* **2001**, *311*, 183.
(53) Kussell, E.; Shimada, J.; Shakhnovich, E. I. *Proteins* **2003**, *52*, 303.

(54) Kikuchi, A.; Nose, T. *Polymer* **1996**, *37*, 5889.
(55) Selb, J.; Gallot, Y. *Makromol. Chem.* **1981**, *182*, 1491, 1513, 1775.
(56) Wu, C.; Qiu, X. *Phys. Rev. Lett.* **1998**, *80*, 620.
(57) Smit, B. In *Computer Simulation in Chemical Physics*; Allen, M. P., Tildesley, D. J., Eds.; Kluwer Academic Publishers: New York, 1993; p 461.
(58) Laschewsky, A. *Adv. Polym. Sci.* **1995**, *124*, 1 and references cited therein.
(59) Borisov, O. V.; Halperin, A. *Langmuir* **1995**, *11*, 2911.
(60) Zhou, S. Q.; Chu, B. *Adv. Mater.* **2000**, *12*, 545.
(61) Borisov, O. V.; Halperin, A. *Curr. Opin. Colloid Interface Sci.* **1998**, *3*, 415.
(62) Kreig, A.; Lefebvre, A. A.; Hahn, H.; Balsara, N. P.; Qi, S.; Chakraborty, A. K.; Xenidou, M.; Hadjichristidis, N. *J. Chem. Phys.* **2001**, *115*, 6243.
(63) Ginnel, R. *Association Theory*; Elsevier: Amsterdam, 1979.
(64) Szleifer, I.; Ben-Shaul, A.; Gelbert, W. M. *J. Chem. Phys.* **1987**, *86*, 7094.
(65) Care, C. M.; Dalby, T. *Europhys. Lett.* **1999**, *45*, 38.
(66) Evans, G. T. *J. Chem. Phys.* **1997**, *106*, 9718.

MA0359741