Adv Polym Sci (2006) 195: 1–100
DOI 10.1007/12_049

Published online: 13 December 2005

Computer-Aided Conformation-Dependent Design of Copolymer Sequences

Pavel G. Khalatur[1,2] (✉) · Alexei R. Khokhlov[1,2,3]

[1]Institute of Organoelement Compounds, Russian Academy of Sciences, 117823 Moscow, Russia
khalatur@germany.ru

[2]Department of Polymer Science, University of Ulm, 89069 Ulm, Germany
khalatur@germany.ru

[3]Physics Department, Moscow State University, 119899 Moscow, Russia

Abstract A survey is given of the simulation methods as applied to the design of nontrivial sequences in synthetic copolymers aimed at achieving desired functional properties. We consider a recently developed approach, called conformation-dependent sequence design (CDSD), which is based on the assumption that a copolymer obtained under certain preparation conditions is able to "remember" features of the original conformation in which it was built and to store the corresponding information in the resulting sequence. The emphasis is on copolymer sequences exhibiting large-scale compositional heterogeneities and long-range statistical correlations between monomer units. Several new synthetic strategies and polymerization processes that allow synthesis of copolymers with a broad variation of their sequence distributions are reported. We demonstrate that the CDSD polymer-analogous transformation is a versatile approach allowing various functional copolymers to be obtained. Another synthetic strategy is the CDSD step growth copolymerization which is carried out under special conditions. It includes the intrinsic possibilities of exploiting the heterogeneities of the reaction system to control the chemical microstructure of the synthesized copolymers, making possible new paradigms for synthesis and production of polymeric materials. In both cases, we try to show how the preparation conditions dictate copolymer sequences. Also, we discuss advances that have recently been achieved in the computer simulation and theoretical understanding of designed copolymers in solution and in bulk. The focus is on amphiphilic protein-like copolymers and on hydrophobic polyelectrolytes. Here, we demonstrate how copolymer sequence dictates structure and properties.

Keywords Charged heteropolymers · Copolymers · Phase behavior · Polyamphiphiles · Sequence design · Simulation · Solution properties

Abbreviations

A	amphiphilic group (monomer, monomeric unit)
α^2	chain expansion factor
$[A_M]$	mole fraction of intermolecular aggregates of size M
ATC	adsorption-tuned copolymer
ATRP	atom transfer radical polymerization
b	bond length
$\boldsymbol{C}$	symmetric matrix of direct correlation functions
χ	Flory–Huggins interaction parameter
$\widetilde{\chi}$	effective interaction parameter
C_α^0	bulk concentration of monomer species α
$C_\alpha(\boldsymbol{r})$	instant local concentration of monomer species α
$C_\alpha(z)$	equilibrium concentration profile of monomer species α
$c(r)$	direct site-site pair correlation function

CRP	controlled radical polymerization
DCTT	degenerative chain transfer technique
CDSD	conformation-dependent sequence design
DF	density functional
DFA	detrended fluctuation analysis
ΔG	change in association free energy
D_{L}	block length dispersion
D_λ	dispersion within a sliding window of length λ
$\Delta\mu_{\mathrm{s}}$	solvation free energy
$\Delta(q)$	determinant of matrix integral equation
ΔT^*	change in transition temperature
ε	energy parameter
$\boldsymbol{E}$	unit diagonal matrix
ε^*	critical adsorption energy
$\varepsilon_{\mathrm{PP}}$	attraction energy between hydrophilic (polar) segments
f	fraction of charged monomers
$\boldsymbol{\Phi}$	volume fraction of macromolecules
φ_α	average fraction of monomer species α in copolymer chain
ϕ_α	volume fraction of monomer species α
$\varphi_\alpha^{(i)}$	intrachain composition profile of monomer species α
$\phi_\alpha(\boldsymbol{r})$	volume fraction field of monomer species α
$F_D(\lambda)$	detrended local fluctuations within a window of length λ
$f(\lambda)$	block length distribution function
F_{s}	sequence free energy
γ	solvation parameter
h	Shannon's entropy
H	hydrophobic monomer (monomeric unit)
$\boldsymbol{H}$	symmetric matrix of total site-site correlation functions
HA model	hydrophobic-amphiphilic (side-chain) model
HPE	hydrophobic polyelectrolyte
HP model	hydrophobic-hydrophilic(polar) model
$h(r)$	total site-site pair correlation function
JS	Jensen–Shannon divergence measure
K	association equilibrium constant
k_{B}	Boltzmann constant
ℓ	block length
L	average block length
λ	length of sliding window along copolymer sequence
L_{H}	average length of hydrophobic blocks
L_{P}	average length of hydrophilic (polar) blocks
LRC	long-range correlation
m	average number of copolymer chains per aggregate
MAST	macrophase separation transition
MFPT	mean first passage time
MIST	microphase separation transition
N	total chain length, number of repeat units
N_α	number of repeat units of type α (α = A, B) in the chain
NIPA	poly(*N*-isopropylacrylamide)
n_{J}	number of crosslinks
N_τ	current chain length

Ω_α	volume occupied by monomer species α
$\omega_\alpha(\boldsymbol{r})$	average chemical potential field of monomer species α
ODT	order-disorder transition
P	hydrophilic (polar) monomer (monomeric unit)
PA	hydrophilic-amphiphilic copolymer
$p_\alpha^{(i)}$	probability that monomer α is located at the ith position in the chain
PCF	site-site pair correlation function
PEO	poly(ethylene oxide)
PM	probabilistic model
PMF	potential of mean force
PMMA	poly(methylmethacrylate)
PRISM	polymer-reference-interaction-site model
PrP	prion protein
PS	polystyrene
$P(\sigma,T)$	probability of copolymer/particle complex
q	wave number
$\boldsymbol{q}$	wave vector
q^*	wave number of maximum instability (peak in the structure factor)
$q(\boldsymbol{r},\boldsymbol{s})$	propagator
$q^\dagger(\boldsymbol{r},\boldsymbol{s})$	conjugate propagator
R	size of micelle
ρ	monomer number density
r^*	spatial scale of microdomain structure (domain size)
R_g	radius of gyration
R_g^2	mean-square gyration radius
$R_\mathrm{g}^\mathrm{app}$	apparent radius of gyration
$R_\mathrm{g}(t)$	time-dependent radius of gyration
R_gH^2	partial mean-square gyration radius of hydrophobic monomers
R_gP^2	partial mean-square gyration radius of hydrophilic monomers
$R_{\mathrm{g}\Theta}^2$	mean-square gyration radius of unperturbed chain
RISM	integral equation reference-interaction-site model
RPA	random phase approximation
r_s	distance between nearest adsorption sites
$\boldsymbol{s}$	contour length of chain
σ	monomer size
σ_α	effective size of monomer species α
SASA	solvent-accessible surface area
$S_\alpha(q)$	partial scattering function for monomer species α
SCF	self-consistent-field
SCMF	self-consistent mean-field
SFRP	stable free-radical polymerization
σ_p	size of “parental” particle
T	absolute temperature
τ	reduced temperature
Θ	Flory theta temperature
T^*	temperature of spinodal instability
$\Theta_{\alpha\beta}^{(i)}$	chemical correlator

T_c	critical temperature of counterion condensation
τ_D	characteristic diffusion time
τ_R	reaction time characterizing polymerization rate
τ_{rel}	chain relaxation time
T_s	sequence design temperature
T_s^*	critical sequence design temperature
$u(r)$	site-site potential describing interaction between nonbonded monomer units
ν	probability of location of terminal reactive site in a given volume
$\boldsymbol{W}$	matrix of intramolecular correlation functions
$w(q)$	single-chain form-factor
$w(r)$	intramolecular site-site correlation function
$W(r)$	radial distribution function of monomeric units
$W(r^*)$	distribution of domain sizes
z	counterion charge (valence)

1 Introduction: Two Paradigms in Sequence Design

Copolymers have been studied extensively for several decades, partly because of their industrial and biological importance, and partly because of their interesting and sometimes perplexing properties. Many physical and mechanical properties of copolymers, which comprise two or more covalently bonded sequences of chemically distinct monomer species, depend on both the comonomer composition and the arrangement of these comonomers in the polymer chain. There may be significant differences, for example, between two polymer systems with the same chemical composition, but one of which has the comonomers randomly distributed in the chain while the other has long blocks of each monomer type.

One may say that in many cases, sequence dictates structure and properties. To illustrate this, we will mention only two familiar examples.

Synthetic block copolymers can spontaneously self-assemble into highly ordered patterns of supramolecular structures (condensed modulated phases), showing a surprisingly rich morphological behavior. These modulated phases with length scales on the order of 1 to 10^3 nm can potentially form the basis for various nanotechnology applications, including the design of synthetic hierarchical materials, and may be effectively controlled by changing the lengths of blocks or their distribution along the chain [1].

Typically, proteins fold to organize a very specific globular conformation, known as the protein's native state, which is in general reasonably stable and unique. It is this well-defined three-dimensional conformation of a polypeptide chain that determines the macroscopic properties and function of a protein. The folding mechanism and biological functionality are directly related to the polypeptide sequence; a completely random amino acid sequence is unlikely to form a functional structure. In this view, polypeptide sequence

forces a protein to be more than a collapsed heteropolymer, but rather to assume a highly specific three-dimensional structure. Hence, a fundamental issue is how functional protein sequences, which determine biologically active structures, differ from random sequences. Understanding the relationship between a protein's sequence and its native structure is one of the key problems in modern science [2–4].

In recent years we have seen intense interest in developing new types of functional polymer macromolecules via clever design of sequences of monomeric units in a copolymer chain. Broadly speaking, sequence design may be defined as an approach aimed at finding the optimum sequence that provides desired properties of the resultant polymer. This requires a scoring function which may typically be based on physical principles, knowledge-based approaches, or a specifically designed function. Alas, insofar as the terms "sequence design" and "sequence engineering" imply a rational, planned approach to the creation or modification of copolymer structure and function, both still remain beyond our capabilities in a general way.

There are two main paradigms in the sequence design problem.

In protein science, the *de novo* sequence design problem consists of finding a sequence of amino acids that fold into a *target* globular structure. This problem is sometimes called the inverse protein folding problem. Many current methods for *de novo* protein sequence design consist of numerically mutating a sequence until a maximum stability is achieved for the target structure that is usually considered as a ground state. There are a number of reviews that cover this subject [5–9]. In polymer chemistry and physics, emphasis is on the development of new methods of synthesis, on the control of (co)polymer stereochemistry and architecture, and on the design of high performance polymeric materials tailored for specific uses and properties.

The difference between the two sequence design concepts is related to several essential differences between natural and non-natural copolymers. We mention here only a few of them.

The order of amino acids in a polypeptide chain produced by the synthetic apparatus of the living cell is always the same for a particular protein so that all the protein sequences of a given type are structurally identical copies in every cell in a living organism. We cannot distinguish one individual protein sequence from another. For most synthetic copolymers, produced industrially or synthesized in research labs, the occurrence of a certain degree of sequence disorder is almost inevitable. Therefore, if we are speaking about a synthetic copolymer sequence, we mean, explicitly or implicitly, that averaging over many different sequences has been carried out. For a protein to function, it must be in its highly specific native conformation that is stable only in a narrow temperature region. On the other hand, the properties and functions of synthetic copolymers are not so tightly related to their con-

formation. Moreover, we are mainly interested in the nonunique copolymer conformations. The interior of proteins has the packing density of a molecular crystal while synthetic globules are typically liquid-like. This list is of course far from exhaustive, but should rather be taken as simply a set of examples.

In the present article we will deal mainly with synthetic macromolecules and practically will not touch on biopolymers.

Diblock and repeated-block AB copolymers are the simplest examples of two-letter copolymers made up of two different monomer species, denoted by letters A and B. More sophisticated distributions of chemically different groups along the chain are characteristic of random and random-block copolymers, including uncorrelated or ideal random copolymers and so-called correlated random copolymers. In the former class, the corresponding chemical sequences are uncorrelated and this corresponds to Bernoulli or zeroth-order Markov processes [10]. In the latter class, the correlation in the sequences of both types of A and B segments is defined by means of a first-order Markov process. It is important to emphasize that in both cases the correlations characterizing the distribution of monomeric units along copolymer chains decay exponentially. There are, however, copolymers for which this is not the case. In this review, we will consider just these copolymers, focusing on the computer-aided design of their chemical sequences as well as on the properties of designed polymers.

Although recent years have witnessed an impressive confluence of experiments and statistical theories, presently there is no comprehensive understanding of the interrelation between chemical sequences in synthetic copolymers and the conditions of synthesis. One has merely to glance at recent literature in polymer science and biophysics to realize that the problem of sequence-property relationship is by no means entirely solved. As always, in these circumstances, an alternative to analytical theories is computer simulations, which are designed to obtain a numerical answer without knowledge of an analytical solution.

The computer simulations are likely to be useful in two distinct situations—the first in which numerical data of a specified accuracy are required, possibly for some utilitarian purpose; the second, perhaps more fundamental, in providing guidance to the theoretician's intuition, e.g., by comparing numerical results with those from approximate analytical approaches. As a consequence, the physical content of the model will depend upon the purpose of the calculation. Our attention here will be focused largely on the coarse-grained (lattice and off-lattice) models of polymers. Naturally, these models should reflect those generic properties of polymers that are the result of the chain-like structure of macromolecules.

Apart from the introductory section, the article is subdivided into two major sections: Synthesis and Properties of designed copolymers.

2
New Synthetic Strategies in Sequence Design

2.1
Preliminary Remarks

Today, the majority of all polymeric materials is produced using the free-radical polymerization technique [11–17]. Unfortunately, however, in conventional free-radical copolymerization, control of the incorporation of monomer species into a copolymer chain is practically impossible. Furthermore, in this process, the propagating macroradicals usually attach monomeric units in a random way, governed by the relative reactivities of polymerizing comonomers. This lack of control confines the versatility of the free-radical process, because the microscopic polymer properties, such as chemical composition distribution and tacticity are key parameters that determine the macroscopic behavior of the resultant product.

The absence of control of the incorporation of monomers into the polymeric chain implies that many macroscopic properties cannot be influenced to a large extent. Therefore, in recent years much effort has been directed towards the development of controlled radical polymerization (CRP) methods for the preparation of various copolymers (for a recent review, see [17]).

These methods are based on the idea of establishing equilibrium between the active and dormant species in solution phase. In particular, the methods include three major techniques called stable free-radical polymerization (SFRP), atom transfer radical polymerization (ATRP), and the degenerative chain transfer technique (DCTT) [17]. Although such syntheses pose significant technical problems, these difficulties have all been successively overcome in the last few years. Nevertheless, the procedure of preparation of the resulting copolymers remains somewhat complicated.

On the other hand, it should be realized that radical copolymerization at heterogeneous conditions offers additional unique opportunities not available in homogeneous (solution) copolymerization. These include the intrinsic possibilities of exploiting the heterogeneities of the reaction system to control the chemical microstructure of the synthesized copolymers, making possible new paradigms for synthesis and production of polymeric materials. In this contribution, we discuss some new synthetic strategies, which have been developed in recent years to provide effective control of the chemical sequences.

In a series of publications [18–20], the concept of conformation-dependent sequence design (CDSD) of functional copolymers has been introduced (for recent reviews, see [21–25]). The essence of the proposed approach is based on the assumption that a copolymer obtained under certain preparation conditions is able to remember features of the original ("parent") conformation in which it was built and to store the corresponding information in the resulting chemical sequence. In other words, this concept takes into account

a strong coupling between the conformation and primary structure of copolymers during their synthesis. Ideologically, the approach [18–20] bears some similarities with that proposed earlier in the context of the problems of protein physics [26–30], however, it is aimed at synthetic copolymers rather than biopolymers. The original idea for protein design [26–30] consisted of running through sequences of amino acids to determine which sequence (or sequences) had the lowest energy in a unique (target) conformation. In the Introduction, we stressed the differences between this approach and CDSD.

In polymer chemistry, there are two known CDSD techniques: (i) the chemical modification (polymer-analogous transformation) of homopolymers and (ii) the step-growth copolymerization of monomers with different properties under special conditions. We will address both these techniques.

Polymer simulations are being done on different levels. An atomistic model of a polymer contains all the atoms that are present in the real polymer. Coarse-grained models simplify the problem by combining atoms into effective united atoms. In this way, only significant microscopic information—"essential features" of the real system —is retained. The unit can represent a chemical group of a few atoms, a monomeric unit in a polymer, groups of monomeric units, or chain segments of various lengths. Certainly, whether a coarse-grained description is adequate for understanding a particular polymer system depends very much on the problem being studied. The challenge lies in selecting just the right amount of atomistic detail to build coarse-grained models with maximum generality. To a certain degree, of course, that is precisely what physics is about. Therefore, the question is how can we construct copolymer models that are sufficiently realistic to capture the essential features of real macromolecules yet simple enough to allow large-scale computations of polymer conformation and dynamics?

There are many kinds of polymerizing monomers used to make up copolymers. These differ in physical and chemical properties. One of the most important differences (essential features) is their solubility, that is, how much they like or dislike a solvent, e.g., water. Hence the chemical and atomistic details of different monomeric units may not be necessary to understand the properties of many "two-letter" copolymers. In what follows, we will mainly use the so-called HP model [31]. This two-letter model of a linear hydrophobic/hydrophilic macromolecule reflects the spirit of minimalist models, in that it is simple yet based on a physical principle.

The HP model is a coarse-grained (lattice or off-lattice) polymer model that abstracts from real polymers in two important ways: (i) Instead of modeling the positions of all atoms of the polymer, it models only the backbone structure of the polymer, i.e., one position for each monomeric unit. (ii) Usually, only the hydrophobic interaction between the monomeric units is modeled, therefore the model distinguishes only two kinds of monomeric units, namely hydrophobic (H) and hydrophilic (or polar, P).

2.2
CDSD Via Polymer-Analogous Modification

2.2.1
Protein-like Copolymers: Structure Dictates Sequence

The studies of structures formed by copolymers consisting of two kinds of monomeric units constitute a rather large field of polymer chemistry and physics [32]. The systems that are most extensively studied are block-copolymers (with a block primary structure) and random copolymers (with statistical primary structure). Sometimes, copolymers with some short-range correlations along the chain are also investigated. Such correlations will always show up after the copolymerization process, if the probability of addition of unit A or B to the growing chain depends on the type of unit that was added on the previous polymerization step [17]. The type of primary structure that emerges in this case can be characterized as "random with short-range correlations". On the other hand, globular proteins can also be regarded from a very rough viewpoint as a kind of binary copolymer. Indeed, the most important difference between the monomeric units of globular proteins is that some amino acid residues are hydrophobic, while others are hydrophilic or charged. We can very roughly attribute the index H to the former type of units and the index P to the latter ones [33]. If we then analyze the primary structures of the globular proteins obtained in this way and compare them with the simple primary structure of conventional synthetic copolymers, we should draw the conclusion that protein-originated AB texts are much more informative and specific.

It is generally believed that in globular proteins the hydrophilic P units mainly cover the surface of the globule, giving rise to their stability against intermolecular aggregation, while hydrophobic H units main form the core of the globule [33]. It can be assumed that such a requirement (in the dense globular state the P amino acids should be on the surface and the H amino acids in the core) is rather restrictive, i.e., it is satisfied only for a very small fraction of all possible primary structures. Moreover, since the HP correlations defined in such a way depend on the conformation of the globule as a whole (i.e., on the ternary structure), they should be characterized as long-range ones.

The question is, whether such primary structures can be obtained for binary copolymers, not obligatorily of biological origin. It is easy to do this by computer simulation [18], and much more difficult in real experiments. However, in both cases the corresponding procedure should involve the following stages that are schematically depicted in Fig. 1:

Stage 1. We take a homopolymer coil with excluded volume interactions in a good solvent.

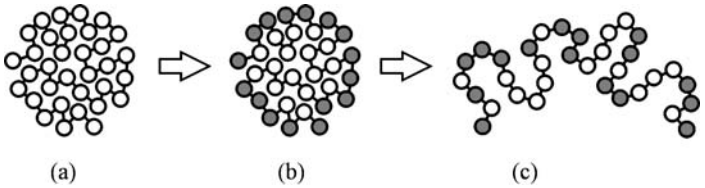

Fig. 1 Schematic representation of the CDSD chemical modification of a homopolymer chain: **a** initial conformation formed by a homopolymer globule; **b** chemical modification (surface "coloring"); **c** resulting copolymer. Modified chain segments are shown in *gray*

Stage 2. Strong attraction between all monomeric units is switched on and a homopolymer globule (parent conformation) is formed (Fig. 1a). Of course, when we are speaking about real experiments, by switching on the attraction we should understand the jump of temperature, addition of poor solvent, etc.

Stage 3. This step is more easily realizable in computer experiments. We should simply consider the "instant photo" of the globule and "color" the units on the surface and in the core in different "colors", i.e., we assign the index P to those units that are on the surface of the globule and call these units hydrophilic and assign the index H to the units in the core of the globule and call these units hydrophobic. Then we fix this primary structure (Fig. 1b). In real experiments the "coloring" of the surface can be done with a chemical reagent Z entering the reaction with monomeric units and converting them from hydrophobic to hydrophilic, $H + Z \rightarrow P$. In chemical language, this is called polymer-analogous transformation. If the amount of reagent is small enough, only surface monomeric units will be contaminated, the core remaining hydrophobic. Another important feature is to have a fast-enough coloring reaction and slow-enough intermolecular aggregation (which will always take place under the conditions when globules are formed). To slow down the aggregation, the neutral thickeners of the aqueous solutions may be used.

Stage 4. This last step is necessary for computer realization. The uniform strong attraction of units should be switched off, and different interaction potentials should be introduced for H and P units.

Initially, the protein-like HP sequences were generated in [18] for the lattice chains of $N = 512$ monomeric units (statistical segments), using for simulations a Monte Carlo method and the lattice bond-fluctuation model [34]. When the chain is a random (quasirandom) heteropolymer, an average over many different sequence distributions must be carried out explicitly to produce the final properties. Therefore, the sequence design scheme was repeated many times, and the results were averaged over different initial configurations.

It turned out that many statistical properties of protein-like and random copolymers with the same HP composition are very different. In order to be able to distinguish whether this difference is due to the special sequence design described above, or just due to the different degree of blockiness, one can introduce for comparison also the random-block primary sequence. The random-block HP copolymers have the same chemical composition and the same average length L of uninterrupted H or P sequence as protein-like copolymers, but in other respects the HP sequence is random. In [18], the distribution of block length λ was taken in the Poisson form: $f(\lambda) = e^{-L}L^{\ell}/\ell!$.

In Fig. 2 we present the typical distributions of H and P monomers along the chain for regular (multiblock) copolymers as well as for purely random, random-block, and protein-like copolymers. From the comparison of the primary structures of the random and protein-like copolymers we can see that the average lengths of H and P blocks in the protein-like copolymer are notably longer. On the other hand, the copolymer with the random-block architecture, having the same average length of H and P sections, L_H and L_P, as for the protein-like copolymer, exhibits a different distribution of these blocks

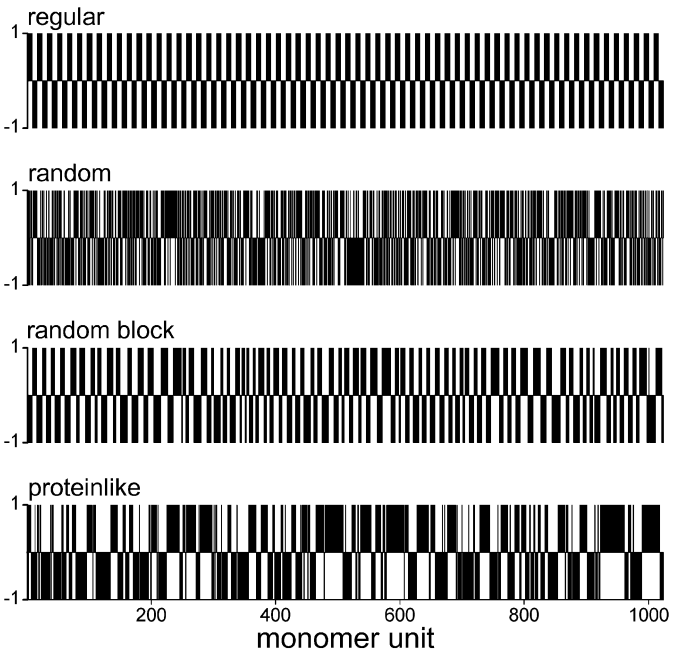

Fig. 2 Typical distributions of H and P monomeric units along the chain for a regular (multiblock) sequence with a fixed block length of 8 units, as well as for a purely random sequence with an average block length $L = 2$, a random-block sequence with $L = 6.4$, and a protein-like sequence with $L = 6.4$. The H units are denoted as +1 and the P units as –1. The sequence length is $N = 1024$

along the chain. A main feature of the protein-like sequences is the presence of rather long uniform H and P sections.

Also, it was found that the coil-to-globule transition for protein-like copolymers, induced by the strong attraction of H segments, occurs at higher temperatures, leads to the formation of denser globules and has faster kinetics than for random and random-block counterparts [18–20]. The reason for this is illustrated in Fig. 3 where the typical snapshots of globules formed by protein-like and random copolymers are shown. One can see that the HP heteropolymer obtained as a result of the simple one-step coloring procedure can self-assemble into a segregated core-shell microstructure thus resembling some of the basic properties of globular proteins. The core of the protein-like globule is much more compact and better formed compared to that observed for random copolymers; it is surrounded by the loops of hydrophilic segments, which stabilize the core.

Apparently, this is due to some *memory effect*: the core, which existed in the parent globular conformation (Fig. 1b), was simply reproduced upon refolding caused by the attraction between H units. One may say that the features of the parent conformation are "inherited" by the protein-like copolymer. Looking at the conformations of Fig. 3, it is natural to argue that protein-like copolymer globules could be soluble in water and thus they are open to

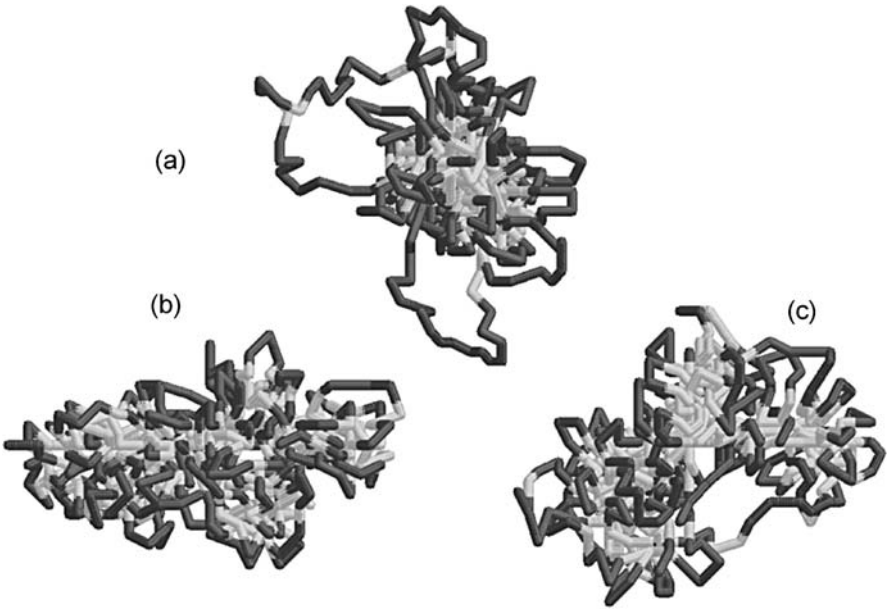

Fig. 3 Typical snapshots of the globular conformations of **a** protein-like, **b** random, and **c** random-block copolymers of the same length ($N = 512$). Hydrophobic segments are shown in *light gray* and hydrophilic segments in *dark gray*. Adapted from [18]

further chemical modification, while random copolymer globules will most probably precipitate. We will address these statements in more detail later.

It is clear that the main reason for the deviation of the properties of protein-like copolymers from those of random copolymers is the special sequence design scheme [18], not just differences in the degree of blockiness.

Govorun et al. [35] have shown, both by exact analytical theory and by computer simulations, that the corresponding chemical sequence is nonalternating and demonstrates the specific long-range correlations (LRCs), which can be described by the statistics of the Lévy-flight type [36]. For such probabilistic processes an observable stochastic variable x exhibits large jumps ("flights"), called "Lévy flights", characterized by a power-law (rather than exponential) probability distribution function, $f(x) \propto x^{-\mu}$ $(1 < \mu < 3)$. The possibility for exact analytical description of sequences resulting from surface coloring (Fig. 1b) comes from the fact that the statistics of polymer chains inside dense globules is Gaussian, i.e., it is described by the ordinary diffusion equation. One has only to worry about correct boundary conditions, and this problem was resolved in [35]. An analogous result was obtained in a later work [37] for the model that takes into account more chemical details.

Experimentally, the conformation-dependent design described above was first realized in a series of papers, where the lyophilization (hydrophilization) of a homopolymer was achieved by the grafting of short poly(ethylene oxide) chains to the surface of globules formed by long poly(*N*-isopropylacrylamide) (NIPA) and glicydil methacrylate chains [38–40]. This synthesis resulted in the formation of a nonrandom copolymer, apparently having a core-shell morphology, in qualitative agreement with simulation data [18–20]. It was shown that grafting to the more compact conformation results in globules that are more stable to precipitation than random grafting to a coil conformation.

2.2.2
Long-Range Correlations and Their Measure

The presence of LRCs in designed sequences is a very important feature. It is easy to understand that these correlations are due to the fact that assigning of the type of chain segments (H or P) under the preparation conditions described above depends on the conformation of the parent globule as a whole (Fig. 1b), not on the conformation of small sections of the initial homopolymer chain. From this viewpoint one may say that such a sequence *encodes* in a two-letter alphabet the spatial (core-shell) structure of a copolymer globule. Obviously, this functional feature can be realized if and only if a general statistical pattern is attributed to the sequence as a whole, and cannot be obtained by the joining of independent statistical patterns of two or more subsequences of smaller lengths.

It may well be that Nature also chose such a path in the evolution of the main biological macromolecules: proteins, DNA, and RNA. These polymers in living systems are responsible for functions, which are incomparably more complex and diverse than the functions that we normally discuss for synthetic polymers. The molecular basis for this ability to perform sophisticated functions is associated with the unique chemical sequences of these biopolymers. In particular, a protein sequence as a whole determines the globular conformation and hence biological function, whereas if this sequence is cut into two pieces, those pieces normally neither correspond to a soluble globule nor have any biological function (Fig. 4). The same is true for statistically complex DNA sequences, which encode in a four-letter alphabet all genetic information and exhibit significant correlation on different scales [41]. All these peculiarities are connected to LRC effects, and the corresponding sequence, which cannot be divided into shorter subsequences with similar statistical patterns and functional features, may be termed an *inseparable sequence*. Such sequence integrity and LRCs are not characteristic of the majority of synthetic linear copolymers, the primary structure of which is chemically

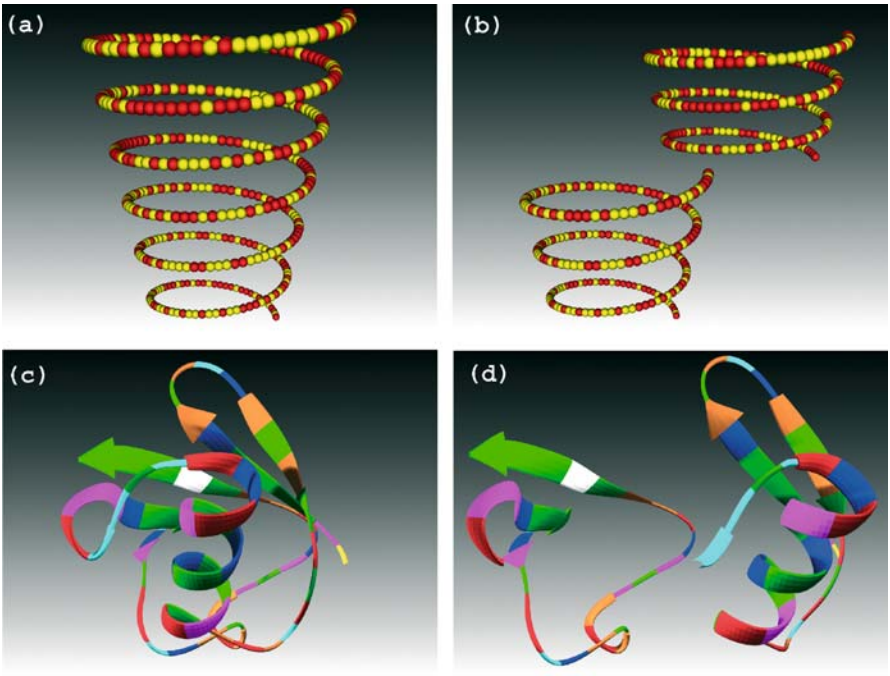

Fig. 4 a,**b** The sequence distribution for rather large sections of a synthetic random copolymer is practically identical to that of the whole polymer. **c**,**d** If a protein sequence is cut into two pieces, those pieces neither correspond to a native globule nor have any biological function

homogeneous on a large scale. Indeed, the sequence distribution for rather large sections of synthetic copolymers is normally practically identical to that of the whole polymer (Fig. 4). Certainly, because of the finite size of the bare globule subjected to CDSD, the longest correlations that can be found in this case are also finite; however, the range of these correlations is much larger than for usual synthetic copolymers obtained, e.g., via radical copolymerization under homogeneous conditions [17]. On the other hand, one may anticipate that if scale-invariant correlations extend along the entire copolymer sequence, that is, the sequence is a "true" fractal object, its sufficiently large parts would have the same statistical pattern and large-scale compositional inhomogeneities. In the present article, we will focus just on such sequences, which show both strong chemical inhomogeneity and LRCs.

How many different, thermodynamically stable structures can be encoded by a long heteropolymer sequence? To answer this question, Fink and Ball [42] have used arguments based on energy fluctuations and information theory. They have found that the maximum number of compact conformations, $p_{\max}$, which are simultaneously thermodynamically stable, depends only on the number of chemically different comonomers in the chain, s, but not on its length N. If a chain is placed on a lattice with an effective coordination number κ (for a cubic lattice, $\kappa \approx 1.85$), the theory predicts: $p_{\max} = \ln(s)/\ln(\kappa)$, that is, for the two-letter alphabet, $p_{\max} \approx 1$. This means that for a binary (two-letter) copolymer, only one nondegenerate ground-state conformation can exist. It is probably not too surprising that binary models are not accurate representations of real proteins. On the other hand they are well suited in polymer chemistry. For a homopolymer, zero conformations are encodable, while for a protein 20-letter amino acid alphabet, the information capacity of the sequence is notably higher: about 5 conformations can be stored. Of course, the theory [42] gives only the number of stable conformations that a heteropolymer can recall, not its general *information complexity* that can be infinitely higher at the $N \to \infty$ limit. Indeed, using the Morse alphabet, we can write all the human history in a message of reasonable length.

For a statistical analysis of copolymer sequences, different mathematical techniques are used. For mathematically oriented researchers, a copolymer sequence might be considered as a string of symbols whose correlation structure can be characterized completely by all possible monomer-monomer correlation functions. Since the correlations at long distances are typically small, it is important to use the best possible estimates to measure the correlations, otherwise the error due to a finite sample size can be as large as the correlation value itself.

To monitor the long-range statistical properties of computer-generated sequences, the method developed by Stanley and co-workers [43, 44] in their search for LRCs in DNA sequences is usually employed. In this approach, each

HP copolymer sequence is transformed into a sequence of + 1 and – 1 symbols (as used in Fig. 2), which are considered as steps of a one-dimensional random walk. Shifting the sliding window of length λ along this sequence step by step, the number of + 1 and – 1 symbols inside the window is counted at each step. This number $\gamma_k(\lambda)$ is a new random variable that depends on the position k of the window along the sequence. The variable $\gamma_k(\lambda)$ has a certain distribution. Its average is determined by the overall sequence composition, and its dispersion is given by $D_\lambda^2 = \langle \gamma^2(\lambda) \rangle - \langle \gamma(\lambda) \rangle^2$, where $\langle \ldots \rangle$ represents the average over all windows of size λ and generated sequences. If the sequence is uncorrelated (normal random walk) or there are only local correlations extending up to a characteristic range (Markov chain), then the value of D_λ scales as $\lambda^{1/2}$ with a window of a sufficiently large λ. A power law $D_\lambda \propto \lambda^\alpha$ with $\alpha > 1/2$ will then manifest the existence of LRCs.

In some cases, because of large fluctuations, conventional scaling analysis cannot be applied reliably to the relatively short sequences generated in typical simulations. To avoid this problem, one can use the so-called detrended fluctuation analysis (DFA), the method specifically adapted to handle problems associated with short nonstationary sequences [43, 44]. This approach leads to a special function $F_D(\lambda)$ that characterizes the detrended local fluctuations within a window of length λ. Generally, $F_D(\lambda)$ shows the same behavior as D_λ. There are many other methods for monitoring LRCs in copolymers [44–47].

The result of calculations [35] averaged over 2000 independent protein-like HP-sequences of $N = 1024$ monomeric units with a 1 : 1 HP composition is presented in Fig. 5. For comparison, the data for two other types of sequences are also shown. One of them is a purely random 1 : 1 sequence; it demonstrates $D_\lambda \propto \lambda^{1/2}$ scaling, as expected. Comparing this curve with

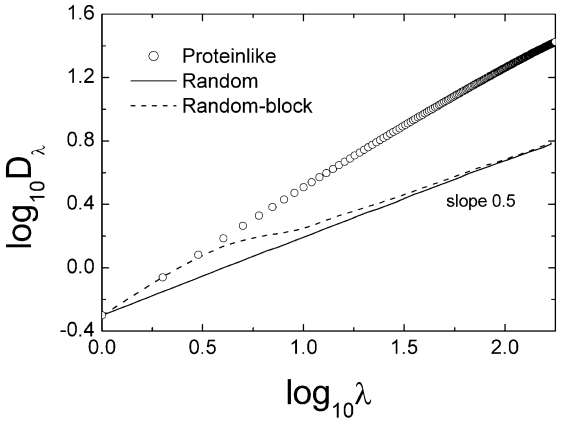

Fig. 5 Dispersion of the number of different monomeric units in a fragment of sequence size λ for protein-like, random, and random-block copolymers. Adapted from [35]

Monte Carlo results we see immediately that the protein-like sequence is not random and well-pronounced correlations do exist in it. Thus, it is interesting to compare the simulation data with the Poisson distribution adjusted to achieve the same 1 : 1 composition and the same degree of blockiness as for a protein-like copolymer. This model sequence exhibits a somewhat more rapid variation of D_λ at small λ, but ultimately the law $D_\lambda \propto \lambda^{1/2}$ is obeyed for large values of λ. Nevertheless, this random-block model is also seen to be unsatisfactory for the statistical behavior of a protein-like sequence throughout the interval of λ examined, $2 < \lambda < 512$. Although the simulation data do not fit accurately to any power law $D_\lambda \propto \lambda^\alpha$, the slope of the observed D_λ dependence corresponds to a significantly larger value than $1/2$, up to about $\alpha = 0.85$, thus indicating pronounced long-range correlations in a protein-like sequence. Thus, the primary structure emerging in the case of protein-like copolymers can be characterized as "quasirandom with long-range correlations". Analytical theory [35] suggests that the Lévy-flight statistics, albeit with a broader crossover region, is expected even if parental (globular) conformation used to generate protein-like macromolecules is not maximally compact, but rather a globule somewhat closer to the Θ-point.

These findings are surprisingly similar to those known for DNA sequences, which appeared as a mosaic of coding and noncoding patches [41, 43, 44]. Indeed, like DNA chains containing coding and noncoding regions, the copolymer under consideration also contains two types of alternating sections forming a certain pattern. It is known that the noncoding regions in DNA do not interrupt the correlation between the coding regions (and vice versa), and the DNA chain is fully correlated throughout its whole length. As a result, the D_λ^2 (or F_D^2) curve does not contain the linear portion $D_\lambda^2 \propto \lambda$. In principle, the same behavior is observed for protein-like sequences [18–20].

The question of whether proteins originate from random sequences of amino acids was addressed in many works. It was demonstrated that protein sequences are not completely random sequences [48]. In particular, the statistical distribution of hydrophobic residues along chains of functional proteins is nonrandom [49]. Furthermore, protein sequences derived from corresponding complete genomes display a distinct multifractal behavior characterized by the so-called generalized Rényi dimensions (instead of a single fractal dimension as in the case of self-similar processes) [50]. It should be kept in mind that sequence correlations in real proteins is a delicate issue which requires a careful analysis.

To end this section, it is worthwhile to note that long-range dependence processes (also called long-memory processes) and their statistics have many areas of application: statistical physics, neuroscience, communication networks, turbulence, hydrology, meteorology, geophysics, finance, econometrics. The literature on the subject is vast (see, e.g., [51]).

2.2.3
Hydrophobic Modification of Hydrophilic Polymers

The segregated core-shell microstructures, consisting of a hydrophobic core surrounded by a hydrophilic shell, are of great practical interest as their mechanical properties are mainly influenced by the core polymer and the chemical properties mainly by the shell monomer units. Recently, a further development of the approach to synthesis of copolymers with a specific primary sequence capable of forming core-shell microstructures has been suggested [25, 52]. The idea is rather similar to the previous approach but has an important difference: instead of the hydrophilization of the globular surface, one can perform a sequential hydrophobization of a hydrophilic polymer chain (poly-P) dissolved in a good solvent, using a low-molecular-weight hydrophobic modifier (H) poorly soluble in this solvent.

In a dilute solution, when the polymer is in a coil state (Fig. 6a), the diffusion of hydrophobic particles into the coil is normally faster than the chemical reaction [53]. In this case, the local concentration of particles H inside the coil is practically the same as in the bulk. Therefore, we expect that at the initial stage, the reaction will lead to a random copolymer: some of the P monomeric units will attach to H reagent and thereby they will acquire amphiphilic (A) properties: $P + H \rightarrow A$ (Fig. 6b). As long as the number of modified A units is not too large, the chain remains in a swollen coil-like conformation (Fig. 6b). However, when this number becomes sufficiently large, the hydrophobically modified polymer segments would tend to form

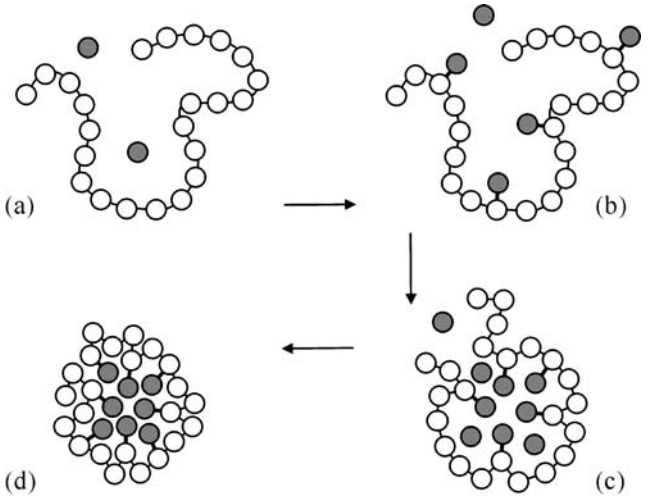

Fig. 6 Schematic representation of the hydrophobic modification of a hydrophilic polymer in a solvent, which is selectively poor for hydrophobic modifier and modified chain segments. The modifying agent and hydrophobic monomers are shown in *gray*

intrachain micelle-like aggregates (Fig. 6c). This is due to the loss in translational entropy of covalently bonded H species after their grafting to the hydrophilic backbone. Structurally, the intramolecular micelles are similar to reverse micelles formed by free low-molecular-weight surfactants, and like ordinary micelles, they should solubilize hydrophobic species.

The presence of the intrachain aggregates can dramatically change the reaction conditions. Because of preferential adsorption, the poorly soluble modifier will diffuse inside the A-rich regions thus leading to spatially inhomogeneous distribution of the concentrations (Fig. 6c). One can expect that further modification of the hydrophilic sections of the polymer chain will occur predominantly within these microglobular regions. In the course of chemical reaction, they would progressively increase in size and then coalesce. As a result, a nonrandom microblock distribution of the chemically modified chain segments would emerge. Finally, we should have hydrophobically modified segments inside rather compact conformation formed by the resulting hydrophilic-amphiphilic (PA) copolymer. In that way, one can expect the formation of core-shell morphologies with inner (poorly soluble) core and outer (well-soluble) hydrophilic cover (Fig. 6d). In principle, the polymer-analogous reaction can be terminated at any desirable time by quenching (e.g., by temperature lowering) or by stopping the supply of reactive compounds.

Given the shortcomings of an approximate analytical treatment and the difficulties with the laboratory measurements, it is conceivable that computer simulations might help greatly in verification of the qualitative arguments presented above.

In the simulation [52], a continuous (off-lattice) model of a polymer and the method of stochastic molecular dynamics were employed. The nonmodified polymer was treated as a flexible chain consisting of N hydrophilic (P) segments (monomeric beads), with adjacent beads connected by a rigid rod of a fixed length b. Each hydrophobic modifier (H) was considered as a single-site particle (monomer). Each amphiphilic group (A) arising after the attachment of H monomer to the hydrophilic backbone was modeled by a two-site "dumbbell" consisting of H and P sites linked by a rigid bond of length b. The hydrophilic sites were connected with each other in a linear fashion and formed the backbone of the hydrophobically modified PA copolymer.

The total potential energy of the system consisting of the solute and solvent molecules was decomposed in the following three parts: (i) the solute-solute potential energy, (ii) the solvent-solvent potential energy, and (iii) the interaction potential energy of the solute and the solvent. An integration over all solvent degrees of freedom leads to the potential of mean force (PMF). This effective solvent-mediated potential can be written as the sum of the solute-solute potential energy and the mean solvation term that describes the solvent-induced effects. The hydrophobic effect is the most important force in stabilizing globular structures. It is believed that the corresponding attrac-

tive interaction is primarily entropy driven and it should be short-range and should depend on a shell of surrounding solvent molecules. Experimental data show that around room temperature for a wide range of different hydrophobic molecules, the hydrophobic interaction energy depends linearly on the burial of solvent-accessible surface area (SASA). Taking into account this fact, to treat the hydrophobic interactions in the simulation [52], an approach based on the SASA model [54] was used.

In this model, the solvation free energy, $\Delta\mu_s$, required to transfer an N-site molecule from vacuum to polar solvent (e.g., in water) is approximated by the sum of linear terms $\Delta\mu_s = \sum_{i=1}^{N} \gamma_i S_i$, where S_i is the solvent-accessible surface area for site i and γ_i is the solvation parameter for the site. In this approximation, $\Delta\mu_s$ can also be considered as the free energy cost for the transfer of a particle from the interior of a hydrophobic cluster to the solvent. The solvation parameter γ is an estimate of the free energy of transfer divided by exposed surface area. For hydrophilic particles, $\gamma \leq 0$. For hydrophobic particles, solvent quality becomes poorer with increasing γ. In reality, when the temperature is fixed, the change in γ can be due to variation of the solvent composition.

The algorithm used in [52] simulated a reaction in a 3D cube utilizing periodic boundary conditions. Initially, an N-unit hydrophilic homopolymer (poly-P) in a coil state and N free hydrophobic H monomers were placed in the cube. For the conditions considered, despite the attraction between H monomers, they were soluble due to their high translational entropy. If a freely diffusing monofunctional H monomer approached within the prescribed distance (reaction radius) to a P bead on the chain, a bond could be formed between the two. This led to the formation of an amphiphilic monomer unit: P + H $\rightarrow$ A. The reaction was considered as a sequence of the alternating steps: grafting of a new H monomer to the chain and the subsequent long relaxation of the chain. The process was terminated when the required number of the hydrophilic monomer units was transformed into the A type. In the study [25, 52], the composition of the resulting copolymer was constrained so that there were 75% hydrophilic and 25% amphiphilic monomer units. As a result, a hydrophilic-amphiphilic PA copolymer with a certain distribution of P and A units along the hydrophilic polymer backbone was obtained. To gain better statistics, ca. 10^3 hydrophobically modified copolymer chains for each set of the parameters were independently generated and then required average characteristics were found.

To give a visual impression of the simulated system, Fig. 7 presents a typical snapshot of an amphiphilic copolymer having a 256-unit hydrophilic backbone with 64 attached hydrophobic side groups. Already from this picture, it is seen that, using the synthetic strategy described above, one can indeed end up with a copolymer having a dense hydrophobic core surrounded by a hydrophilic shell.

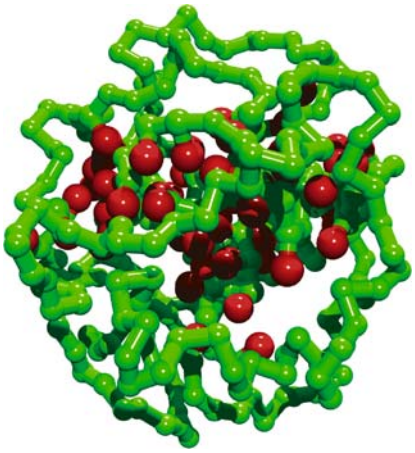

Fig. 7 Typical snapshot of a hydrophobically modified copolymer. Hydrophilic chain segments are shown in *green* and hydrophobic side groups in *red*

From the picture presented in Fig. 7, one can expect that the sequential hydrophobization of a polymer coil should lead to a copolymer with a nonrandom sequence distribution. This is indeed the case. As an example, let us consider the average number fractions of blocks consisting of ℓ neighboring amphiphilic monomers, $f_A(\ell)$, occurring in a copolymer chain. Some results are shown in Fig. 8 on a semilogarithmic scale.

We expect that for a random distribution of monomers A incorporated into the polymer chain, the function $f_A(\ell)$ should decay exponentially with increasing ℓ. In fact, such a behavior is observed for the copolymer modified in a solvent, which is good for the low-molecular-weight modifier, that is, at $\gamma = 0$. When the solvation parameter γ is increased and the solvent becomes selectively poorer for the modifying agent, the values of $f_A(\ell)$ are skewed toward A sections of greater length, which implies a copolymer with blocky tendencies [10]. A further worsening of the solvent quality leads to a strong deviation from the exponential decay of $f_A(\ell)$ thus indicating a nonrandom copolymer sequence distribution.

Thus, copolymers of the same composition can have qualitatively different sequence distributions depending on the solvent in which the chemical transformation is performed. In a solvent selectively poor for modifying agent, hydrophobically-modified copolymers were found to have the sequence distribution with LRCs, whereas in a nonselective (good) solvent, the reaction always leads to the formation of random (Bernoullian) copolymers [52]. In the former case, the chemical microstructure cannot be described by any Markov process, contrary to the majority of conventional synthetic copolymers [10].

In general, the microsegregated structures observed for hydrophobically modified polymers (Fig. 7) are similar to core-shell globules obtained via

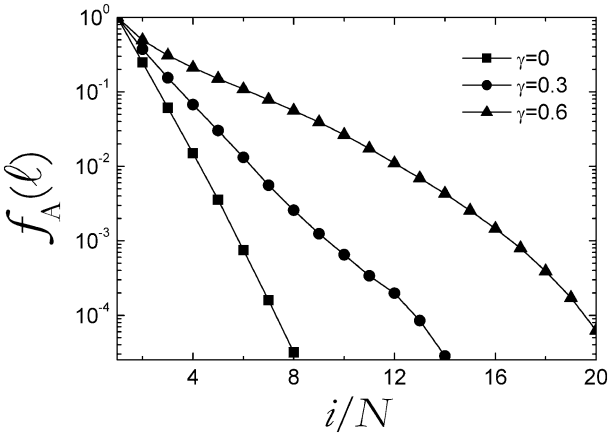

Fig. 8 Average number fraction of blocks consisting of ℓ neighboring amphiphilic monomers occurring in a 256-unit copolymer chain hydrophobically modified in a solvent having different selectivity for modifying agent. Adapted from [25]

the coloring procedure (Fig. 3). However, it should be kept in mind that the experimental realization of the hydrophilic modification of the surface of a hydrophobic globule was shown to be rather unreliable, because of the difficulty of stabilizing dense globules in the solution for the time sufficient to implement a polymer-analogous transformation [23]. On the other hand, the simulations [25, 52] show that the method based on the hydrophobic modification of soluble polymers should be quite universal and robust.

2.2.4 Adsorption-Tuned Copolymers

The idea of conformation-dependent sequence design via polymer-analogous transformation can be generalized in many respects [23]. Indeed, a special chemical sequence can be obtained not only from a globular conformation; any specific polymer chain conformation can play the role of parent.

The simplest example of this kind is connected to the conformation of a homopolymer partly adsorbed onto a flat substrate (Fig. 9). Let us assume that the chain segments being in direct contact with the surface in some typical instant conformation (Fig. 9a) are chemically modified (Fig. 9b). This can take place when the surface catalyzes some chemical transformation of the adsorbed segments. One can expect that after desorption (Fig. 9c), such a copolymer will have special functional properties: it will be "tuned to adsorption".

Following this line, Zheligovskaya et al. [55] compared the adsorption properties of copolymers with special "adsorption-tuned" primary structures (adsorption-tuned copolymers, ATCs) with those of truly random copoly-

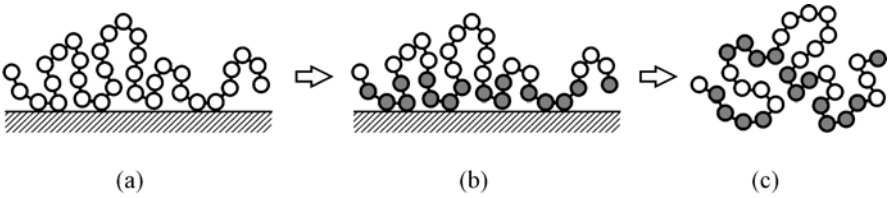

Fig. 9 Schematic representation of the sequence design procedure leading to an adsorption-tuned copolymer: **a** initial partly adsorbed homopolymer, **b** chemical modification of adsorbed chain segments, **c** resulting copolymer. Modified segments are shown in *gray*

mers and random-block copolymers. Monte Carlo simulations revealed that specific features of the ATC primary structure promoted the adsorption of ATC chains, in comparison with their random and random-block counterparts under the same conditions. In other words, the resulting copolymer sequence "memorizes" the original state of the adsorbed homopolymer chain. Its statistical properties exhibit LRCs of the Lévy-flight type similar to those found for copolymers obtained via coloring of a homopolymer globule [56].

Recently, Velichko et al. [57] suggested the model of a so-called *molecular dispenser*. This idea is a further development in the direction of conformation-dependent sequence design. Namely, they considered the conformation of a homopolymer chain adsorbed on a spherical colloidal particle (Fig. 10) and performed design of sequence for this state of macromolecule. The motivation behind this design procedure is that if we eliminate the parent colloidal particle after the design is completed (e.g., by etching), the resulting copolymer will be hopefully tuned to selectively adsorb another colloidal particle of a parental size σ_p. For instance, if such a copolymer is exposed to a polydisperse colloidal solution of particles of different size, it will selectively choose to form a complex with the particle having the same radius as that in parental conditions. That is why such a macromolecular object can be called a molecular dispenser.

It should be noted that the development of such polymer systems is stimulated by existing experimental works. In particular, the experimental methods of preparation of nanometer-sized hollow-sphere structures have been suggested [58–63] because of their possible usage for encapsulation of molecules or colloidal particles. The preparation of hollow-sphere structures, generally, is based on self-assembling properties of block copolymers in a selective solvent, i.e., on the formation of polymer micelles with a nanometer-sized diameter. Further cross-linking of the shell of the micelle and photodegradation [64] of the core part produce nanometer-sized hollow cross-linked micelles.

The sequence design procedure proposed in [57] can be described in more detail as follows.

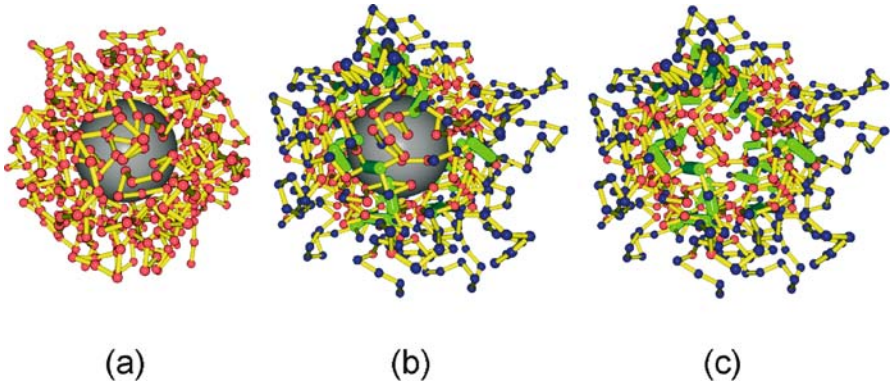

Fig. 10 Stages of preparation of a copolymer envelope: **a** adsorption of homopolymer chain on a colloidal particle; **b** "coloring" of the polymer chain (*blue* corresponds to chemically modified monomer units and *red* to adsorbed units) and introduction of crosslinks (shown as *green sticks*) to stabilize the hollow-spherical structure; **c** elimination of the core particle. Adapted from [57]

First, we consider a homopolymer chain attracting to a colloidal nanoparticle. Such a chain forms an adsorbed complex with the particle (its typical conformation is shown in Fig. 10a). Only part of the chain segments are in direct contact with the colloidal particle, while other segments form flower-like loops. Then we color the segments in the loops in "blue", while the segments near the colloidal particle remain "red", i.e., they are attracting to this particle (Fig. 10b). If the sequence design is stopped at this stage, the pronounced selectivity of the complex formation with another particle of parental size σ_p is not reached [57]. However, if additional crosslinks are introduced between adsorbed (red) units, thus fixing the cage structure of the central cavity (Fig. 10b), the macromolecule emerging after elimination of the colloidal particle (Fig. 10c) does indeed show the features of a molecular dispenser. This sequence design scheme was realized in the computer simulations [57]. We will address the results of this work in Sect. 3.5.2.

2.2.5 Design as a Simulation of Evolutionary Process

The concept of evolution of primary sequences of biopolymers has attracted great interest from biologists, chemists and physicists for a long time [65–68]. As has been discussed, it is natural to expect that the content of information in the sequences of biopolymers (proteins, DNA, RNA) is relatively high in comparison with random sequences where it should be almost zero [69]. Presumably, the information complexity of early ancestors of present-day biopolymers has been increased in the course of molecular evolution when the

copolymer sequences became more and more complicated [66]. The study of various possibilities of this evolution of copolymer sequences is just the area where the evolution concept can be used in the context of polymer science.

It is worthwhile to note that since the information content of a sequence can be represented as a mathematically defined quantity, the whole process of evolution of biopolymer sequences can be specified in exact mathematical terms. The formulated fundamental problem is extremely difficult because of the absence of direct information on the early prebiological evolution. Therefore, of particular interest are "toy models" of evolution of sequences that show different possibilities for appearance of statistical complexity and of long-range correlations in the sequences.

It is clear that information complexity cannot emerge just as a result of random mutations. Some coupling of mutations to other factors is necessary. It is most straightforward and natural to introduce the interrelation of mutations and conformations, i.e., to consider conformation-dependent sequence design in the context of evolution.

Evolutionary computation approaches are optimization methods. They are conveniently presented using the metaphor of natural evolution: a randomly initialized population of individuals evolves following a crude parody of the Darwinian principle of the survival of the fittest. New individuals are generated using simulated evolutionary operations such as mutations. The probability of survival of the newly generated solutions depends on their fitness (how well they perform with respect to the optimization problem at hand): the "best" are kept with a high probability, the "worst" are rapidly discarded.

In the literature, some computer models describing the evolution of copolymer sequences have been proposed [26, 28]. Most of them are based on a stochastic Monte Carlo optimization principle (Metropolis scheme) and aimed at the problems of protein physics. Such optimization algorithms start with arbitrary sequences and proceed by making random substitutions biased to minimize relative potential energy of the initial sequence and/or to maximize the folding rate of the target structure.

It should be emphasized that the problem, which we address here, is somewhat different from that usually discussed in the context of protein physics [5–9]. We are not aiming at the search for a unique three-dimensional (native) conformation with a fast folding rate. On the contrary, we are interested in a state with a large entropy. In general, our aim is to learn whether it is possible to make with synthetic copolymers a step along the same line as molecular evolution.

Ascending and descending branches of sequence evolution. The aim of the study [70] was to introduce explicitly the concept of sequence evolution into the CDSD scheme.

Using a molecular-dynamics-based algorithm, the conformation-dependent evolution of model HP copolymer sequences was simulated [70]. The se-

quence evolution mechanism involved the generation of an initial protein-like sequence by inspecting of a homopolymer globule and by attributing the H-type to the monomeric units in the core of this globule and the P-type to the units on the surface of the globule. The resulting copolymer was then transferred to a coil conformation and then refolded due to the strong attraction between the H units. The HP sequence was then further modified, depending on the position of a monomer in the core or on the surface of a newly formed globule. Such modifications, leading to a change in the primary HP sequence, are repeated many times (ca. 10^3). With this evolutionary process, which can be called "repeated coloring", structures and sequences are formed self-consistently.

A 128-unit flexible-chain heteropolymer with a HP composition fixed at 1 : 1 was simulated for the condition when hydrophobic H monomers strongly attract each other, thus stabilizing a dense globular core, whereas the attraction energy ε_{PP} between hydrophilic P monomers was considered as a parameter (the interaction between H and P monomers is given by $\varepsilon_{HP} = \sqrt{\varepsilon_{HH}\varepsilon_{PP}}$). For this model system, various conformation-dependent and sequence-dependent properties, including information-theoretic-based quantities, can be calculated.

Depending on the attraction energy between polar segments ε_{PP}, it is possible to find two regimes (branches) of evolution (regimes I and II) [70]. If ε_{PP} is smaller than some crossover energy ε^*_{PP} (regime I), the evolution can lead to a second-order-like transition in sequence space from the sequences with a protein-like primary structure capable of forming a core-shell globule to the degenerated (nonprotein-like) sequences having long uniform H and P blocks. This transition is also accompanied by strong changes in the conformational properties of the copolymer.

Figure 11a shows the mean square gyration radius, R^2_g, plotted vs. ε_{PP}. As seen, R^2_g is a weakly decreasing function of ε_{PP} in the range $\varepsilon_{PP} > \varepsilon^*_{PP}$ and demonstrates a rapid growth when ε_{PP} decreases and becomes less than ε^*_{PP}. The critical value ε^*_{PP} is found to be smaller than the critical energy at which the coil-to-globule transition takes place in a homopolymer chain of the same length.

The degenerated primary structure looks like a di- or triblock sequence ("core-tail" or "tadpole-like" conformation).

Thus, when the attraction between hydrophilic segments is not sufficiently strong, we deal with the *descending branch* of the evolution, which leads to nonprotein-like sequences having low information content and low complexity. On the other hand, in the second regime (at $\varepsilon_{PP} \geq \varepsilon^*_{PP}$) the complexity of protein-like structures is found to increase and therefore we have the *ascending branch* of the evolution.

Information complexity of copolymer sequences. A common approach to the analysis of the complexity of a system is to use concepts from information theory and information-theoretic-based techniques.

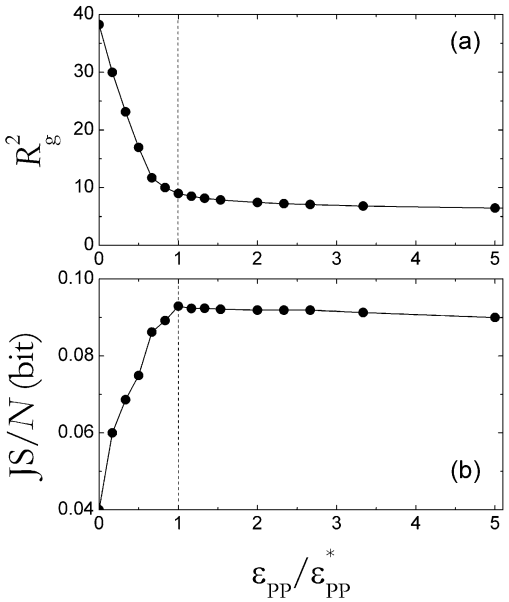

Fig. 11 **a** Mean square gyration radius and **b** Jensen–Shannon divergence measure as a function of the attraction energy ε_{PP} between hydrophilic segments, after the sequence evolution procedure. The characteristic energy of H – H interactions is fixed at $\varepsilon_{HH} = 2k_B T$, thus stabilizing a dense globular core. Adapted from [70]

What is complexity? There is no good general definition of complexity, though there are many. Intuitively, complexity lies somewhere between order and disorder, between regularity and randomness, between perfect crystal and gas. Complexity has been measured by logical depth, metric entropy, information content (Shannon's entropy), fluctuation complexity, and many other techniques; some of them are discussed below. These measures are well suited to specific physical or chemical applications, but none describe the general features of complexity. Obviously, the lack of a definition of complexity does not prevent researchers from using the term.

In general, the aim here is to find a measure capable of indicating how far copolymer sequences generated during the evolutionary process differ from each other and from random or trivial (degenerate) sequences. It turned out that the usual measures of the degree of complexity (based, e.g., on Shannon's entropy and related characteristics) are nonadequate [70]. To overcome this problem, it was proposed to use the so-called Jensen–Shannon (*JS*) divergence measure [70]. Let us explain how it can be defined.

Let $\mathbf{S} = \{s_1, ..., s_N\}$ be a sequence of N symbols. For two subsequences $\mathbf{S}_1 = \{s_1, ..., s_n\}$ and $\mathbf{S}_2 = \{s_{n+1}, ..., s_N\}$ of lengths n and $N - n$, the difference between the corresponding discrete probability distributions $f_1(s_1, ..., s_n)$ and

$f_2(s_{n+1}, \ldots, s_N)$ is quantified by the Jensen–Shannon divergence

$$JS(\mathbf{S}_1, \mathbf{S}_2)/N = h(\mathbf{S}) - \left[\frac{n}{N}h(\mathbf{S}_1) + \frac{N-n}{N}h(\mathbf{S}_2)\right], \quad (1)$$

where $\mathbf{S} = \mathbf{S}_1 \oplus \mathbf{S}_2$ (concatenation) and $h(\mathbf{S})$ is Shannon's entropy of the empirical probability distribution obtained from block frequencies in the corresponding subsequence. Of course, Shannon's entropy depends on the definition of a set of words in the sequence. For two-letter HP copolymers, one can adopt the following set of words (uniform blocks): H, HH, HHH, ..., P, PP, PPP, ...; that is, the word (block) is defined by its length ℓ and type. In this case, Shannon's entropy per monomer can be written as

$$h = -\frac{N_w}{2N}\sum_{\ell}\left[f_H(\ell)\log_2 f_H(\ell) + f_P(\ell)\log_2 f_P(\ell)\right], \quad (2)$$

where $f_H(\ell)$ and $f_P(\ell)$ are the frequencies of words of length ℓ composed of letters H and P, respectively, and N_w is the total number of words.

The Jensen–Shannon divergence *JS* is zero for subsequences with the same statistical characteristics; it takes higher values for increasing differences between the statistical patterns in the subsequences, and reaches its maximum value for a certain set of distributions. In particular, both random and any regular (multiblock) copolymers of infinite length show $JS = 0$. We normally expect that a completely random sequence or a sequence with long uniform blocks contain less information than a sequence containing many different blocks (words) of medium length. Of course, only using sequence analysis, we cannot unambiguously distinguish between what might be called quantity and quality of information.

Using the Jensen–Shannon divergence, *JS*, as a measure of complexity for the generated sequences, one can obtain an interesting result (see Fig. 11b). The most important feature is that the *JS* value is a *nonmonotonous* function of ε_{PP}, whereas Shannon's entropy, Shannon's index, and many other sequence-dependent parameters change always *gradually* [70].

For the sequences generated in the evolutionary process described above, it was shown that at $\varepsilon_{PP} \geq \varepsilon^*_{PP}$ (regime II) the degree of complexity, as measured by *JS* divergence, can be considerably higher as compared to that observed for regime I, at $\varepsilon_{PP} < \varepsilon^*_{PP}$. The complexity slightly increases with ε_{PP} decreasing, reaches its maximum just on the boundary of regimes I and II, and then sharply drops (Fig. 11b). Therefore, in regime II, the evolution preserved the copolymer sequence of high complexity, whereas in regime I, the information content of the sequence has degenerated in the course of evolution.

Simultaneous evolution of sequences and conformations. In the work [71], the simultaneous evolution of sequences and conformations was studied. This design procedure leads to the final state that depends on the set of interaction parameters and on the rearrangements both in conformational space

and in sequence space. These rearrangements are characterized by the usual thermodynamic temperature, T, for conformational space as well as effective sequence rearrangement temperature, T_s. Namely, after a certain number of Monte Carlo steps in conformational space (in the course of this process the system equilibrates at temperature T) the possibility of mutation of two randomly chosen monomeric units is tried: monomeric unit H converts into P and vice versa. If this leads to a decrease in the energy of a protein-like globule, this move is accepted. If the globular energy is increasing by the amount ΔE, this move is accepted with the probability defined by $\exp(-\Delta E/k_B T_s)$.

In general, we can define evolution of sequences for any value of T_s. However, the simulation for three characteristic cases can be most easily understood.

(i) The inequality $T \ll T_s$ means that all the moves in sequence space are accepted independent of conformation. This corresponds to random mutations, and the final sequence (after long evolution) will be that of a random HP copolymer (no information complexity).

(ii) The inequality $T \gg T_s$ means that only the moves leading to a decrease in the globular energy E are accepted. Such evolution should lead to a sequence corresponding to a minimum of globular free energy in conformational space. For the case of the absence of any attractive interactions between P units, it was shown [70] that the final sequence after evolution should have a hydrophobic core with very few hydrophilic (or polar) loops and a long hydrophilic tail. This sequence is close to that of a HP diblock copolymer, and should not exhibit any information complexity, as has been stated above.

(iii) The case $T = T_s$ corresponds to an annealed HP sequence. This case is equivalent to the situation when monomeric unit H can be converted into P by attaching some ligand L: $P \rightleftarrows H + L$. We assume that the number of ligands is fixed to maintain 1 : 1 HP composition, however, they can choose which monomeric unit to bind. This defines, in particular, the chain sequence. Annealed HP sequences in the context of polymer globules were first considered by Grosberg [72].

When T_s does not correspond to any of the characteristic cases described above, the evolution of sequences coupled with conformations can be still defined in the same way. One has only to remember that if $T \neq T_s$ and both T and T_s are finite, there is a flow of heat between conformational space and sequence space, so that full thermodynamic equilibrium is impossible. Still, we can be in a stationary regime corresponding to the sequences tending to a certain fixed point, and possessing (or not possessing) information complexity.

The simulations and theoretical arguments [71] predict that at high T_s, the sequence free energy F_s dominates and the sequence tends to be completely random, corresponding to the minimum of F_s. At low T_s, the evolution selects those sequences, which correspond to the low value of the conformation

free energy (structure of the core-tail type). In the intermediate regime, the conformational thermodynamic force and the sequence contribution (evolution pressure) interplay. Therefore, the formation of nontrivial structures and sequences is possible. Even in the absence of attraction between P units the final sequences remain protein-like (i.e., the final copolymer has the core-shell structure of a globule) and therefore maintains certain information complexity.

2.3 CDSD Via Copolymerization

2.3.1 Conditions for CDSD

Polymer growth is an example of a hybrid stochastic process, which involves both discrete and continuous random variables; the position of the polymer being continuous, while the number of monomers in the growing polymer is discrete. In this section, we will discuss just those processes.

The CDSD copolymerization essentially differs in principles and in results from the conformation-dependent polymer-analogous transformation [25]. Such a process can be considered as a variant of template polymerization (also called molecular imprinting) based on the noncovalent binding of polymerizing monomers to a template.

The essence of this technique consists of the copolymerization of monomers differing in their affinity to the template and therefore differently distributed in the reaction system. In contrast to conventional types of template polymerization [73, 74], in the CDSD copolymerization, all the monomers are bifunctional and form linear polymers, not cross-linked ones. The sequence of the segments in the resulting molecularly imprinted copolymer is determined by the template-controlled conformation of the propagating macroradical [24, 25].

Thus, the CDSD copolymerization regime is possible only in reaction systems with a strongly inhomogeneous spatial distribution of monomer concentrations [24, 25]. As an origin of inhomogeneities, one can mention interfaces, nanoparticles, molecular and supramolecular aggregates (e.g., for macroradicals capable of forming globular conformations), and so forth. They should selectively adsorb one of the comonomers (Fig. 12). Moreover, the spatial scale of the concentration inhomogeneities should be comparable to the size of the growing macroradical. Also, the polymerizing monomers should retain their properties after they are incorporated into a polymer chain. In the case of a copolymerization near an interface, one of the types of monomers and chain segments should be preferentially distributed near the adsorbing surface, whereas other monomers and chain segments should be preferentially located in the solution. As a result, the chemical sequence and

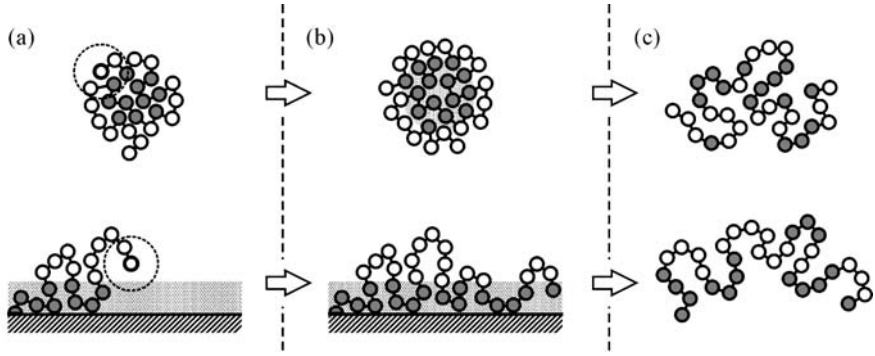

Fig. 12 Schematic representation of the CDSD copolymerization process in the cases when one of the comonomers is selectively absorbed by a polymer globule (*top*) or adsorbed on a surface (*bottom*). **a** Growing chains during the copolymerization (reaction zone around the growing chain end is marked with a *dashed line*); resulting copolymers in **b** globular (adsorbed) and **c** coil-like states. Regions where absorbed (adsorbed) monomers dominate are shown in *gray*

conformation of the growing macroradical become mutually dependent. Such an interdependence determines the further way of chain growth through the comonomer concentrations in a small reaction volume (see Fig. 12).

Finally, the rate of copolymerization should be slow enough. This guarantees that, during the chemical reaction, the equilibrium concentration fields remain approximately constant, and the growing chain has an equilibrium conformation between successive attachments of the monomers. Therefore, the CDSD regime is realized when

$$\max\{\tau_D, \tau_{rel}\} < \tau_R \,. \tag{3}$$

where τ_R is the reaction time characterizing the polymerization rate, τ_D is the characteristic diffusion time for monomers, and τ_{rel} is the chain relaxation time that depends on the current chain length N_τ. For an ideal chain, the Rouse relaxation time scales with N as $\tau_{rel} \sim N^2$ [75]. It is clear that Eq. 3 corresponds to a kinetically controlled regime.

2.3.2
Copolymerization with Simultaneous Globule Formation

Let us assume that we are performing a radical copolymerization of moderately hydrophobic (H) and hydrophilic (P) monomers in an aqueous medium. The conditions for copolymerization (e.g., the temperature and solvent composition) should be chosen in such a way that when the emerging chain is long enough it can form a globule. As long as the current chain length N_τ is not too large, the growing hydrophobic-hydrophilic (HP) macroradical remains in a coil-like conformation. However, when its length becomes suffi-

ciently large, the chain tends to form a two-layer globule with a hydrophobic core and a polar envelope. As has been noted in Sect. 2.2.3, the origin of this effect is connected to the loss in translational entropy of covalently bonded hydrophobic species after their incorporation into the growing chain.

The presence of a hydrophobic-hydrophilic interface can dramatically change the reaction conditions. The hydrophobic core will selectively absorb hydrophobic species from the solution (Fig. 12), and this will result in a redistribution of monomer concentrations between the core and bulk solution. Because the probability of attachment for each comonomer is determined by its concentration in a relatively small reaction volume near an active chain end, the active center inside the hydrophobic core will mainly attach more hydrophobic species; on the other hand, when the active center is located on the globule surface, it will mainly attach polar (soluble) monomers. In this way, the two-layer globule will grow, retaining its core-shell structure with a predominantly hydrophobic core and a hydrophilic outer envelope (see Fig. 12).

Theoretically, chain conformations and chemical sequences obtained via conformation-dependent copolymerization in a selective solvent were studied by Berezkin et al. [76–78]. It was shown that the corresponding polymerization process could not be interpreted in terms of classical kinetic models allowing for only short-range effects [53]. Using Monte Carlo and molecular dynamics simulation techniques, the process of irreversible radical copolymerization of hydrophobic and hydrophilic monomers, which led to globule formation, was studied [76–78]. The polymerization was modeled as a step-by-step chemical reaction of the addition of H and P monomers to the growing copolymer chain, under the assumption that a depolymerization reaction was not allowed. To simplify the analysis, it was assumed that the copolymerization is ideal with equal reactivity ratios. The type of attached monomers and the probabilities of their addition were determined from the average concentrations of the reactive monomers in a reaction volume V_τ around the moving active end of the macroradical of a given length N_τ during a reaction time τ_R. It was also assumed that the reaction bath is sufficiently large and the reactive species are sufficiently dilute so that there is significant time for the growing macroradicals to propagate independently. The preferential sorption of hydrophobic monomers in the core of the arising globule was explicitly taken into account.

The simulations [76–78] show that, using the conformation-dependent polymerization mechanism, one can indeed end up with a copolymer having a dense hydrophobic core surrounded by a hydrophilic shell. Figure 13 presents the typical radial distribution $W(r)$ of hydrophobic and hydrophilic units (with respect to the center of the globule) calculated for a copolymer chain of 512 units with a 1 : 1 HP composition. In Fig. 14, we show a typical snapshot. These findings explicitly prove the core-shell microstructure of the globule obtained via irreversible copolymerization in the solvent, which is moderately poor for hydrophobic species. In general, the same microsegre-

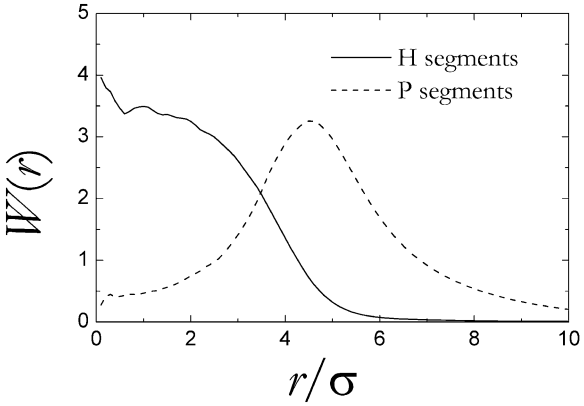

Fig. 13 Radial distribution of hydrophobic and hydrophilic segments (with respect to the center of the globule) for a copolymer chain of 512 segments with a 1 : 1 HP composition. σ is the size of the chain segment. Adapted from [76]

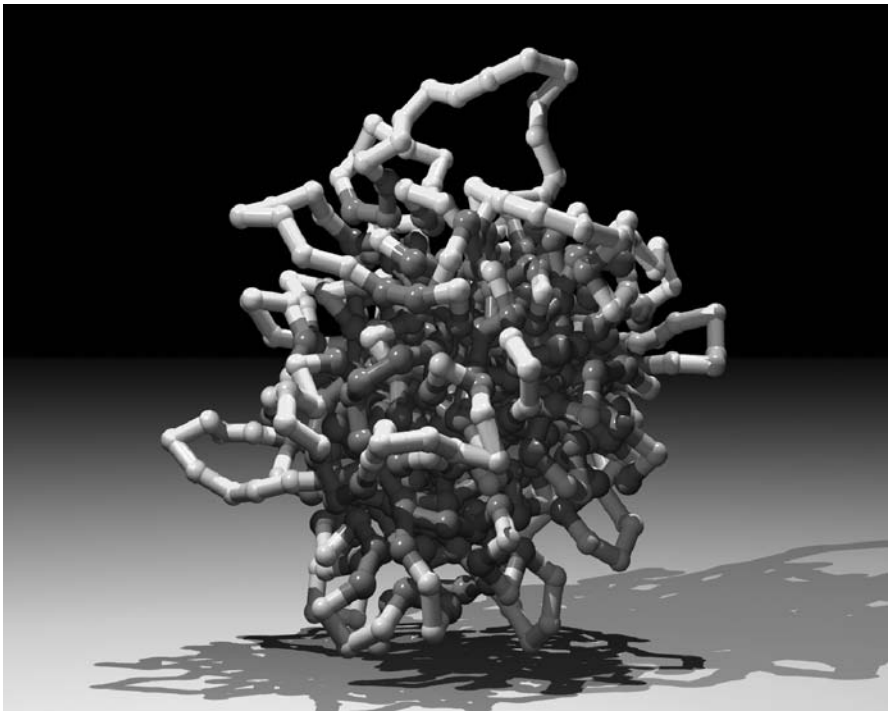

Fig. 14 Typical core-shell microsegregated structure that is obtained via conformation-dependent copolymerization with simultaneous globule formation in a selective solvent [76]. Hydrophobic chain segments are shown in *dark gray* and hydrophilic segments are shown in *light gray*

gated structure is observed for core-shell globules obtained via the coloring procedure [18–20].

The average chemical composition of the resulting HP polymer is determined by the bulk concentrations of the monomers, C_H^0 and C_P^0, and by the solubility of hydrophobic monomers in a globular core. For approximately equal contents of H and P segments in the growing copolymer, the solution should contain an excess of P monomers. In general, the chemical composition of the synthesized copolymer strongly deviates from the statistical one. Thus, the change in the solution concentration of polymerizing monomers allows the copolymer composition to vary over a wide range. Although the reactivities of both monomers are equal [76–78], their copolymerization leads to a copolymer that is preferentially enriched with hydrophobic segments. Therefore, a strong deviation from the principle of equal reactivity of Flory [79] is observed for the CDSD polymerization regime.

The type of monomer attached to the growth center during the simulation under kinetic control (at large τ_R values, see Eq. 3) is determined by the conformation and primary structure of the growing chain as a whole, not only by the local concentration of reactive monomers near the active end of the macroradical. As a result of such cooperativity, the formation of sequences with specific LRCs of the Lévy-flight type was observed [76–78].

Because of the specific mechanism of the chain propagation, the synthesized copolymers can have a gradient primary structure; that is, the HP composition can change along the growing chain during copolymerization [77]. This interesting result follows from the analysis of the distribution of monomers along the growing chain. Very short chains do not contain a sufficiently large number of hydrophobic units to form a core-shell conformation, and the growth of these chains is similar to that observed for random free radical copolymerization [17]. The composition of these chains is also close to the monomer composition in solution. Then, as the macroradical becomes longer and the number of connected H units increases, a dense hydrophobic core can be formed. This core absorbs H monomers, and their fraction in the resulting copolymer increases. In this case, the probabilities of monomer addition are also not constant because of the changes in the ratio between the volumes of the hydrophobic core and the polar shell during the chain growth. Therefore, the gradient primary structure is formed because of a change in the chain conformation and a continuous redistribution of comonomers between the globule and the solution in the course of the polymerization; in this way, a compositional drift is produced along the chain.

Experimentally, the described synthetic strategy was first realized by Lozinsky et al. [80, 81], who studied the redox-initiated free-radical copolymerization of thermosensitive *N*-vinylcaprolactam with hydrophilic *N*-vinylimidazole at different temperatures. These and other experimental studies [82–84] showed the universality of this approach of obtaining copolymers capable of forming nanostructures with a core-shell morphology.

2.3.3 Emulsion Copolymerization

The emulsion polymerization methodology is one of the most important commercial processes. The simplest system for an emulsion (co)polymerization consists of water-insoluble monomers, surfactants in a concentration above the CMC, and a water-soluble initiator, when all these species are placed in water. Initially, the system is emulsified. This results in the formation of thermodynamically stable micelles or microemulsions built up from monomer (nano)droplets stabilized by surfactants. The system is then agitated, e.g., by heating it. This leads to thermal decomposition of the initiator and free-radical polymerization starts [85]. Here, we will consider a somewhat unusual scenario, when a surfactant behaves as a polymerizing comonomer [25, 86].

In the numerical Monte Carlo simulations [86], polymerizing monomers were modeled as hydrophobic/hydrophobic (HH) and hydrophobic/hydrophilic (HP) "dumbbells" consisting of H(P) beads linked by rigid bonds of a fixed length. It is assumed that the monomers are placed in a polar solvent and form a mixed spherical micelle, which can be divided into two regions, Ω_{HH} and Ω_{HP}, initially filled with hydrophobic (HH) and amphiphilic (HP) monomers (Fig. 15a). Also, it is assumed that there is a strong incompatibility of H and P sites (the strong segregation regime). This results in a well-defined

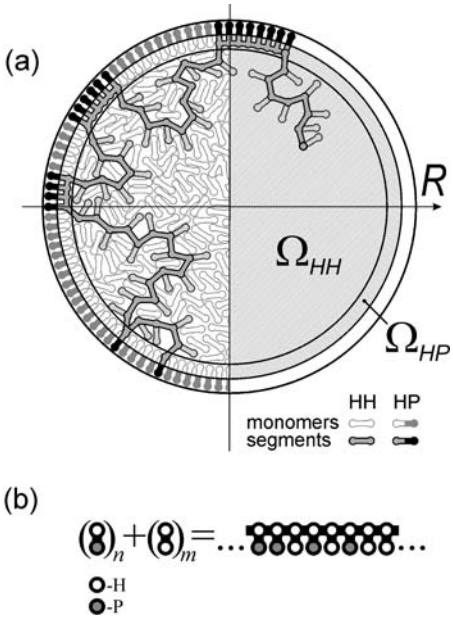

Fig. 15 **a** Schematic representation of the distribution of hydrophobic and amphiphilic monomers polymerizing in micelle and **b** the reaction between the monomers

interface between the Ω_{HH} and Ω_{HP} regions. The polymerization process is modeled as a step-by-step irreversible chemical reaction between H beads, which leads to an amphiphilic copolymer having a hydrophobic backbone with attached hydrophobic and hydrophilic side groups (Fig. 15b). The distribution of the side groups along the hydrophobic backbone defines the primary structure of the resulting amphiphilic copolymer. It is clear that in the model, the copolymer composition should depend on the size of the micelle R and on the ratio of the volumes Ω_{HH}/Ω_{HP}. The maximum length of the chain is limited and defined as $N = 2\pi\rho R^3/3$, where ρ is the total number density of H and P sites in the micelle. Therefore, by changing the micelle size, one can control both chemical composition and chain length simultaneously.

The type of attached monomeric units is determined by the average concentrations of reactive monomers in a reaction volume swept out by the moving active end of the macroradical of a given length N_τ during time τ_R. These concentrations are related to the corresponding instant local concentrations, $C_{HP}(\boldsymbol{r})$ and $C_{HH}(\boldsymbol{r})$, in the elementary volume $d^3\boldsymbol{r}$ and to the probability of finding the active end in this volume, $\nu(\boldsymbol{r})$. The probability $p_{\alpha\beta}^{(i)}$ that monomeric unit $\alpha\beta$ ($\alpha\beta$ = HP or HH) is located at the ith position from the beginning of a growing macromolecule is given by [86]

$$p_{\alpha\beta}^{(i)} = \frac{\int\limits_{\Omega_{\alpha\beta}} C_{\alpha\beta}(\boldsymbol{r})\nu(\boldsymbol{r})\mathrm{d}^3\boldsymbol{r}}{\int\limits_{\Omega_{HP}} C_{HP}(\boldsymbol{r})\nu(\boldsymbol{r})\mathrm{d}^3\boldsymbol{r} + \int\limits_{\Omega_{HH}} C_{HH}(\boldsymbol{r})\nu(\boldsymbol{r})\mathrm{d}^3\boldsymbol{r}} = \frac{\nu_{\alpha\beta}C_{\alpha\beta}}{\nu_{HP}C_{HP} + \nu_{HH}C_{HH}}, \quad (4)$$

where $C_{\alpha\beta}$ is the time-dependent average concentration of monomer $\alpha\beta$ in the corresponding volume $\Omega_{\alpha\beta}$ and $\nu_{\alpha\beta}$ is the probability averaged over chain conformations, which define the location of the terminal free-radical reactive site in this volume.

Generally speaking, the values ν_{HP} and ν_{HH} depend on the conformation of the macroradical as a whole. In the first approximation, however, they can be considered as constant, $\overline{\nu}_{HP}$ and $\overline{\nu}_{HH}$. In this case, the polymerization process is described on the basis of the first-order reaction kinetic equations. In particular, one can define the following parameter characterizing the composition change along the chain [86]

$$\eta = \frac{\overline{\nu}_{HP}}{\overline{\nu}_{HH}} \frac{C_{HP}^{(0)} n_{HH}^{(0)}}{C_{HH}^{(0)} n_{HP}^{(0)}}, \quad (5)$$

where $C_{HP}^{(0)}$ and $C_{HH}^{(0)}$ are the initial monomer concentrations and $n_{HP}^{(0)}$ and $n_{HH}^{(0)}$ denote the initial numbers of the corresponding monomers in the micelle. It turns out that if $\eta > 1$, then the growing chain end attaches mainly hydrophobic monomers; if $\eta < 1$, it is preferentially enriched with hydrophilic monomers. In both cases, the resulting copolymer has a gradient structure. At $\eta = 1$, different monomers are distributed randomly along the chain. Obvi-

ously, the parameter η depends on the micelle size and therefore on N. Also, it can be changed when some amount of chemically inert species (e.g., solvent molecules) is incorporated into the micelle interior.

Let us consider the intramolecular composition profile $\varphi_{\mathrm{HP}}^{(i)} = 2p_{\mathrm{HP}}^{(i)} - 1$ calculated in [86]. The function $\varphi_{\mathrm{HP}}^{(i)}$ characterizes intramolecular chemical inhomogeneity along the chain. The present definition of $\varphi_{\mathrm{HP}}^{(i)}$ assumes that the HP-type segments are coded by symbol + 1, whereas symbol – 1 is assigned to the HH-type segments. For an ideal random copolymer in which chemically different segments follow each other in a statistically random fashion, the $p_{\mathrm{HP}}^{(i)}$ function should coincide with the average fraction φ_{HP} of HP segments for any i. For a random-block copolymer, the fraction of one component averaged over many generated chains should also be uniform along the chain.

Figure 16 shows $\varphi_{\mathrm{HP}}^{(i)}$ as a function of i/N for a few values of N. It is seen that depending on the chain length (and, therefore, on the micelle size), the resulting copolymer can have different gradient primary structures.

Also, in [86], the block length distribution functions $f_{\mathrm{HP}}(\ell)$ and $f_{\mathrm{HH}}(\ell)$ were calculated. It was found that the $f_{\mathrm{HP}}(\ell)$ function decays exponentially with increasing block length ℓ, $f_{\mathrm{HP}}(\ell) \propto \exp(-\ell/L_{\mathrm{HP}})$, where L_{HP} is the average length of HP blocks. On the other hand, the distribution $f_{\mathrm{HH}}(\ell)$ exhibits power-law decay $f_{\mathrm{HH}}(\ell) \propto \ell^{-\alpha}$ for not too large ℓ. The exponent α estimated from the simulation data is close to 2. Such a behavior indicates the existence of LRCs. However, for sufficiently long blocks, $f_{\mathrm{HH}}(\ell)$ shows an exponential decay that is related to the finite size of the micelle.

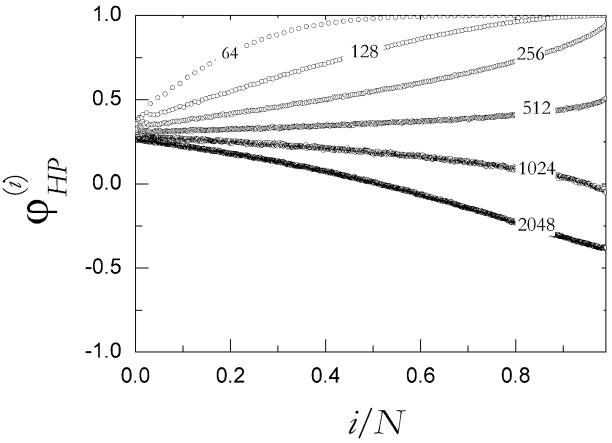

Fig. 16 Intramolecular composition profile presented as a function of monomer number i/N for hydrophobic/amphiphilic copolymer chains obtained via emulsion copolymerization, at a few chain lengths N indicated near the curves. Adapted from [25]

2.3.4
Copolymerization Near a Selectively Adsorbing Surface

Recently, a Monte Carlo simulation technique has been used to study two-letter quasirandom copolymers, which were generated via surface-induced computer-aided sequence design [87, 88]. This approach represents an irreversible radical copolymerization of selectively adsorbed A and B monomers with different affinity to a chemically homogeneous impenetrable surface, allowing for a strong short-range monomer(A)-surface attraction. Thereby, one of the types of monomers and chain segments are preferentially distributed near the adsorbing surface while other monomers and chain segments are preferentially located in the solution. As a result of such concentration inhomogeneities, the chemical sequence and conformation of the growing macroradical become mutually dependent, which corresponds to the CDSD regime [24, 25]. In the simulations, probabilities of A and B monomer additions to the growing polymer were determined by combination of the equilibrium monomer concentration profiles, $C_A(z)$ and $C_B(z)$, normal to the $z = 0$ plane and by the active chain end location. To describe the chain propagation theoretically, a simple analytical model based on stochastic processes and probabilistic statistics was also introduced and investigated in detail [88]. This probabilistic model (PM) provides a close approximation to the simulation data and explains a number of statistical properties of copolymer sequences. The calculations revealed the following conclusions.

In the case of strong adsorption of A species, the average fraction of B units φ_B in the resulting copolymer increases with their average bulk concentration C_B^0, thereby leading to a decrease in the number of strongly adsorbed segments. At some critical value C_B^*, about half of the chain segments are in the adsorbed state, and the average AB composition of synthesized copolymer is close to equimolar ($\varphi_A \approx \varphi_B \approx 1/2$). At $C_B^0 < C_B^*$, the growing chain is enriched with A monomers; when $C_B^0 > C_B^*$ the growing chain end deeply penetrates into the bulk, and the B segments prevail in the generated copolymer. Thus, the change in the solution concentration of unadsorbed (or weakly adsorbed) reactive monomers allows the chemical composition to be varied over a wide range. Figure 17 presents a typical snapshot picture of the synthesized 1024-segment copolymer chain with $\varphi_A \approx 0.5$.

The chain propagation near the adsorbing surface proceeds as a randomly alternating growth, leading to a copolymer with power-law long-range correlations in the distribution of different segments along the chain. Moreover, the statistical properties of the copolymer sequences correspond to those of a one-dimensional *fractal* object with scale-invariant correlations [88].

Figure 18 presents the results of the statistical analysis for the sequences having approximately 1 : 1 AB composition. Also, Fig. 18 demonstrates the fluctuation function $F_D(\lambda)$ predicted by the probabilistic model [88] for $N = 1024$ and $\varphi_A = 1/2$. We see that designed sequences do not correspond

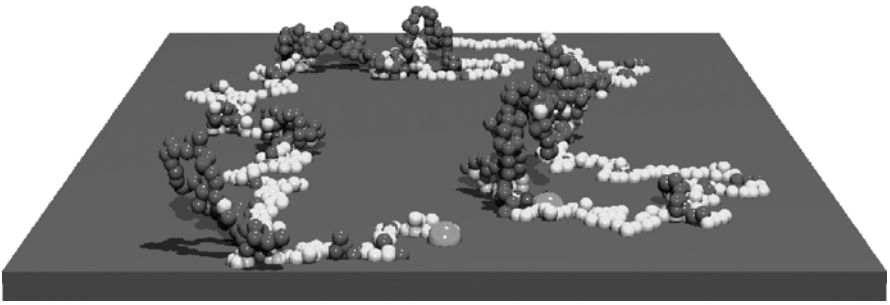

Fig. 17 Snapshot of the synthesized 1024-unit copolymer chain with $\varphi_A \approx 0.5$. The *light gray* and *dark gray spheres* show adsorbed (A) and unadsorbed (B) segments, respectively. The place of connection of the chain with the surface ("initiator") and the chain end are depicted as *larger spheres*

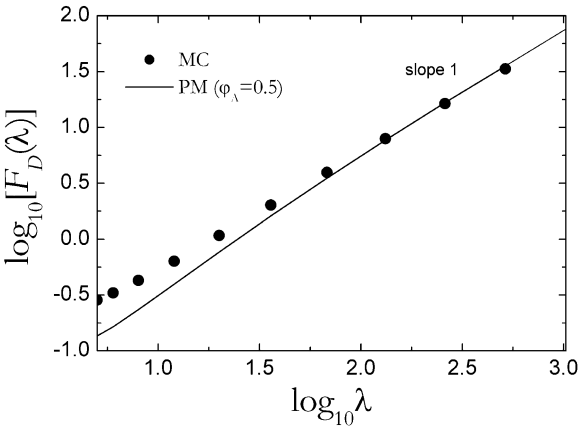

Fig. 18 Detrended fluctuation function for the copolymer sequences of length $N = 1024$, which were generated via surface-induced computer-aided design. *Solid line* shows the analytical (PM) result at $\varphi_A = 0.5$. Adapted from [88]

to random and random-block statistics, and strong correlations exist in these sequences. For sufficiently large λ, very good agreement between the Monte Carlo simulation and the analytical PM result is observed [88]. In both cases, the long-range correlations persist up to the windows with length close to N. Moreover, the correlations turn out to be more pronounced as λ is increased: the dependence of $\log[F_D(\lambda)]$ on $\log\lambda$ becomes a nearly linear function whose slope approaches unity with increasing λ.

From the facts presented above, it is evident that the copolymer sequences discussed here are correlated throughout their whole length. Also, it was found that any sufficiently large part of the averaged sequence has practically the same correlation properties as the entire sequence. This means that the generated sequences show scale invariance, a feature typical of *fractal structures*.

To gain some further insight into the discussed problem, the block length distribution functions $f_A(\ell)$ and $f_B(\ell)$ were calculated, using both Monte Carlo simulations and the analytical probabilistic model [88].

In the $N \to \infty$ limit, the distribution over strongly adsorbed blocks $f_A(\ell)$ is described by the following conditional probability distribution function

$$f_A(\ell) \equiv P_A(\ell|\ell > 0) = (1 - p_{AA})p_{AA}^{\ell-1}, \tag{6}$$

where $\ell = 1, 2, \ldots, p_{AA} = p_A/[1 - (1 - p_A)(1 - p_B)]$ is the conditional probabilities that the $(i + 1)$th segment in the growing chain is of type A when the ith segment is also of type A, p_A is the probability that the copolymer is lengthened by addition of the next monomeric unit of the type A to a terminal free-radical reactive site, and p_B defines the probability that at each ith step the growing chain goes further from the adsorbing surface.

One may say that the conditional probability p_{AA} defines the strength of persistent correlations in the sequence. If the persistent correlations are extremely strong ($p_{AA} \to 1$), then for the average length of A blocks, we have: $L_A \to \infty$. In the polymerization process, such a situation is realized when the bulk concentration of B monomers approaches zero. If $p_{AA} = 1/2$, one arrives at the known trivial result for the Bernoullian statistics without correlations that corresponds to a random copolymer [10]. In this case, the probability p_A of each segment to be of type A is constant throughout the whole sequence. This means, for example, that the conditional probability p_{AA} is equal to p_A. The average fraction of type A segments φ_A in a purely random sequence is equal to the probability p_A. In particular, for a purely random sequence with $p_A = 1/2$ we have: $L_A = L_B = 2$. Such a copolymer is obtained in the course of solution copolymerization of A and B monomers having identical solubility and reactivity. Finally, at $p_{AA} = 0$ we deal with a regular sequence with alternating distribution of A and B segments in the sequence for which $L_A = L_B = 1$.

For copolymer chains of a finite length N, the probability distribution over the length of blocks is given by [88]

$$f(\ell \le N) = \begin{cases} f(\ell), & \ell = 1, 2, \ldots, N - 1 \\ 1 - \sum_{n=1}^{\ell-1} f(n), & \ell = N \end{cases} \tag{7}$$

where $f(\ell)$ is defined from Eq. 6 for blocks A, whereas for blocks B, one has:

$$f_B(\ell = 2m) = \frac{p_{AA}}{1 - p_{AA}} \sum_{n_j} \prod_{j=1}^{m} R_{n_j}, \qquad \sum_{j=1}^{m} n_j = m \tag{8}$$

$$f_B(\ell = 2m + 1) = 0, \qquad m = 1, 2, \ldots \tag{9}$$

with

$$R_n = (1 - p_{AA})f(k = 2n|k > 0), \qquad n = 1, 2, \ldots, m \tag{10}$$

and $R_0 = 1$. Here the conditional probability $f(k = 2n|k > 0)$ is given by

$$f(k = 2n|k > 0) = \begin{cases} 0\,, & n = 2m - 1 \\ \phi_{m-1} - \phi_m\,, & n = 2m \end{cases}, \quad m = 1, 2, \ldots\,, \tag{11}$$

where

$$\phi_m = \sum_{j=1}^{m} b_{j,m}\,, \qquad m = 1, 2, \ldots \tag{12}$$

$$b_{j,m} = \frac{j}{j-1}\frac{m-j+1}{m+j} b_{j-1,m}\,, \qquad j = 2, \ldots, m \tag{13}$$

$$b_{1,m} = \frac{a_m}{m+1} \tag{14}$$

$$a_m = (1 - \frac{1}{2m}) a_{m-1}\,, \qquad m = 2, 3, \ldots \tag{15}$$

and $a_1 = 1$.

One can show that the block length distribution function $f_B(\ell)$ is characterized for asymptotically large ℓ by the power-law decay of its density $f_B(\ell) \propto \ell^{-\alpha}$ with the exponent $\alpha = 3/2$ [88]. The exponent α estimated from the simulation data ($\alpha \approx 1.6$) is, up to the "experimental" uncertainty, quite close to that predicted by the exact analytical model. As has been noted above, such a power-law decay is a characteristic of the Lévy probabilistic processes [36].

It was found that the resulting copolymer has a specific quasigradient primary structure. Figure 19 demonstrates a typical composition profile calculated as a function of i/N for the sequences, which have approximately 1 : 1 AB composition ($\varphi_A \approx \varphi_B \approx 0.5$). First, we observe rather good (almost quantitative) agreement between Monte Carlo and PM results. Second, both profiles show a monotonous decrease with the segment position *i*. Therefore, we deal with a copolymer whose primary structure is similar to that known for "tapered" or gradient copolymers exhibiting strong composition inhomogeneity along their chain [17]. It follows from the simulations and the analytical data that the gradient extends along the entire chain for any chain length. Therefore, the copolymer synthesized in this way shows compositional scale invariance. An inhomogeneous spatial distribution of polymerizing monomers and composition constraints in the resulting copolymer are major reasons behind the specific chain growth and the global statistical nature of sequences at large scales.

It is known that the gradient (tapered) nature of copolymers, which can be synthesized in free-radical polymerization processes, is due to a drift in the free monomer composition during solution polymerization [17]. Such copolymers can be considered as a special type of block copolymers in which the composition of one component varies along the chain. With a decreasing difference in the monomer reactivity rations, the formation of gradient sta-

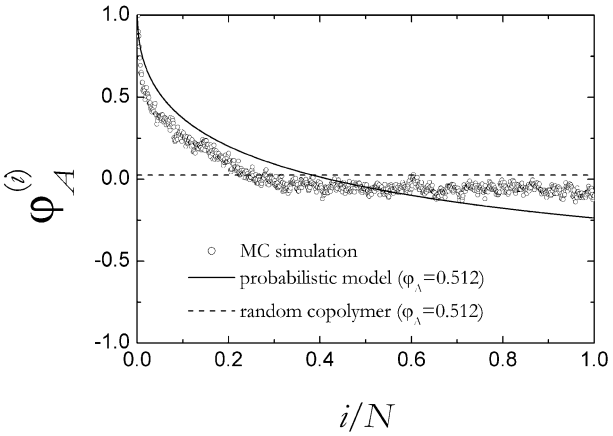

Fig. 19 Intramolecular composition profiles presented as a function of i/N for the sequences having approximately 1 : 1 AB composition ($\varphi_A \approx \varphi_B \approx 0.5$): *circles*, simulation data for the copolymer chains synthesized near a selectively adsorbing surface; *solid line*, probabilistic model, *dashed line*, random sequence. Adapted from [88]

tistical copolymers rather than gradient (tapered) block copolymers occurs. However, in the model polymerization process considered in [87, 88], all the polymerizing species had the same reactivities, and the monomer concentrations remained unchanged during synthesis. Therefore, the changes in the probabilities of the addition of components to the growing macroradical are due to the evolution of its chemical composition; this result is typical for the CDSD regime [24, 25].

2.3.5 Copolymerization Near a Patterned Surface

Because the spatial scale of the monomer concentration gradient in the reaction system discussed in the previous sections was comparable with the macroradical size, the chemical inhomogeneity has been observed for the generated copolymer sequence as a whole. To limit and control the size of the gradient regions, the spatial scale of the concentration inhomogeneities in the reaction system apparently should be limited. The easiest way to achieve this is to perform the copolymerization near a solid patterned surface with discrete adsorption sites. In this case, each of the sites is a small independent source of concentration disturbances lying at a certain distance from other sites. The surfaces with a regular distribution of adsorption sites are of most interest because they can allow a fine control of the primary structure of the copolymers obtained on the basis of the CDSD technique.

Another motivation for this work is the development of copolymers that are tuned to a certain surface, that is, copolymers that have a “memory”

of the preparation conditions and are able to reproduce their specific conformation in the vicinity of the surface with predefined chemical heterogeneity. It is thought that copolymers designed in this way have potential for the recognition of patterned substrates through the formation of stable adsorption complexes with planar or spherical substrates composed of two chemically distinct sites, one of which has a preferential affinity for one of the comonomers. Obviously, these copolymers could have enormous potential in molecular technology and biotechnology. In particular, interfacial molecular recognition is ubiquitous and essential in life processes. Examples include enzyme-substrate binding, transmembrane signaling processes, antigen-antibody and protein-receptor interactions [89, 90]. One of the main motivations for studying the pattern recognition mechanism is also related to the design and synthesis of a new generation of copolymers, which carry a target pattern encoded in their sequence distribution and thereby possess the surface recognition ability [91, 92]. Such copolymers can be viewed as macromolecules with pattern-matched sequences.

It should be emphasized that pattern recognition by synthetic flexible polymers is different from biospecific molecular recognition, which is observed, e.g., for folded globular proteins with a specific ternary structure. Indeed, enzyme-substrate or protein-receptor binding can be explained on the basis of the so-called "lock-key" recognition mechanism [90], whereas the recognition by the flexible coil-like polymers is more likely to be described as "pattern-induced conformational fitting" accompanied by strong conformational changes of the chain [93]. In this latter case, the polymer should adjust its conformation such that it touches predominantly the most attractive surface sites [93]. Therefore, the pattern recognition ability of target-imprinted chemical sequences can in principle be optimized via conformation-dependent sequence design [24].

A Monte Carlo simulation of irreversible template copolymerization near a chemically heterogeneous surface with a regular (hexagonal) distribution of discrete adsorption sites was performed in the papers [94, 95]. The sites could selectively adsorb from solution one of the two polymerizing monomers and the corresponding chain segments. The focus of this study was on the influence of the polymerization rate, adsorption energy ε, and distance between adsorption sites r_s on the chain conformation and chemical sequence of the resulting AB copolymers and, specifically, on the coupling between polymerization and selective adsorption.

Under the preparation conditions corresponding to the CDSD regime, the formation of quasiregular copolymers with a blocky primary structure was observed [94]. In such copolymers, there are two types of alternating sections. One of them contains randomly distributed A and B segments. The second one consists mainly of strongly adsorbed A segments. The average length of the random sections is proportional to the distance separating the nearest neighbor adsorption sites r_s. The average length of the A-rich sections is de-

termined by the adsorption capacity of the adsorption sites. Therefore, by varying the interaction parameters and the distribution of adsorption sites on the substrate, one can design and synthesize copolymers with different surface-induced chemical sequences in a controlled fashion.

In particular, the variation of the strength of the effective monomer(A)-substrate interaction ε allows us to determine the asymptotic regimes corresponding to random copolymerization or CDSD. The average intramolecular chemical composition φ_A emerging in the simulation [94] is shown as a function of ε in Fig. 20 for the case in which the solution concentrations of A and B monomers are equal, $C_A^0 = C_B^0$. One can conclude that for the weak adsorption regime, $\varepsilon/k_BT \lesssim 6$, random copolymerization dominates ($\varphi_A \approx 0.5$); in the region $\varepsilon/k_BT \gtrsim 15$, CDSD dominates. The threshold value of the energy parameter ε when random copolymerization and CDSD have about the same probability is $\varepsilon/k_BT \approx 8$ for a given choice of the polymerization parameters and simulation model.

To analyze the correlations in the copolymer sequences, one can use the so-called two-point "chemical correlators" [53]:

$$\Theta_{\alpha\beta}^{(i)} = n_{\alpha\beta}^{(i)} / \sum_{\alpha,\beta} n_{\alpha\beta}^{(i)}, \tag{16}$$

where $n_{\alpha\beta}^{(i)}$ is the number of $\alpha\beta$ (AA, BB, and AB) pairs and i is the chemical distance between these pairs along the chain. Chemical correlator $\Theta_{\alpha\beta}^{(i)}$ is the probability of finding a pair of α and β segments (α,β = A,B) among all the possible pairs separated by i segments along the chain in a given copolymer sequence. Some of the results of the calculation [94] are shown

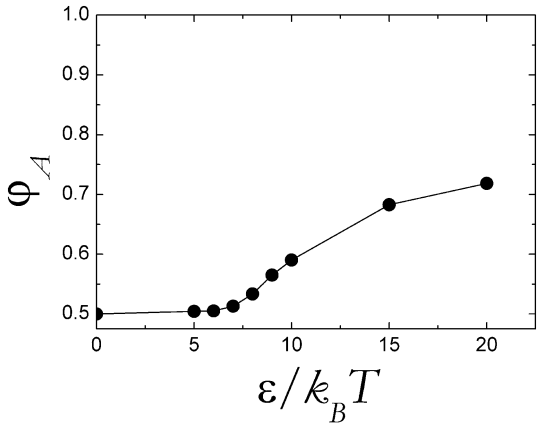

Fig. 20 Average chemical composition as a function of the adsorption energy parameter for the case when the solution concentrations of A and B monomers are equal. Adapted from [94]

in Fig. 21 for a few different values of ε. The dependences of $\Theta_{AA}^{(i)}$ on i indicate that quasiregular copolymers are synthesized in the CDSD regime when the adsorption interaction is sufficiently strong. Indeed, we see that these copolymers are characterized by a periodical variation of the composition along the chain. For the system simulated, the period of this variation is about 20 segments. Therefore, the formation of copolymers with blocky primary structures can be observed.

The following mechanism of the formation of quasiregular copolymers was suggested [94]. Chain growth begins near an adsorption site at which the concentration of strongly adsorbed A monomers is high. These monomers are attached preferably to the end of the growing macroradical and form an initial chain section including mainly A segments. This section gradually covers the nearest adsorption site. Because of the limited adsorption capacity, the screened adsorption site looses the ability to attract the active end group of the macroradical. After that, the reaction volume moves into solution away from the substrate in which random copolymerization occurs. The random chain section grows until it reaches the nearest free adsorption site. Then, the formation of a new adsorbed section begins again. Such cycles are repeated many times. As a result, the formation of a quasiregular copolymer with alternating A-rich and random AB sections is observed.

Thus, the copolymer sequence consists of repeating blocks. Each of these blocks is a short gradient sequence formed by strongly adsorbed sections and random weakly adsorbed sections. The average length of the adsorbed section is almost constant and is related to the adsorption capacity of the adsorption center. On the other hand, the average length of the random section depends on its conformation between the adsorption centers, and in principle, it can vary over a wide range. It should be emphasized that the

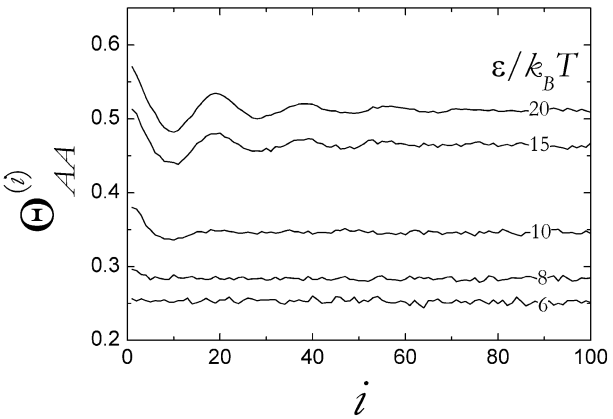

Fig. 21 Chemical correlators for AA diades for different values of the adsorption energy parameter ε, at $N = 512$. Adapted from [94]

quasiregular copolymer is formed only when the random bridges connecting neighboring adsorption sites are strongly stretched and the variation of their length is sufficiently small. The strongly stretched regime leads to the periodical composition variations. Figure 21 shows that this regime is realized for $\varepsilon/k_{\mathrm{B}}T \geq 15$.

The distribution of B blocks, which are included mostly in nonadsorbed chain sections, decays exponentially and thus should obey Bernoullian statistics that correspond to a zeroth-order Markov process [10]. The average length of such blocks is close to 2, that is, the same as that of a random copolymer. In the case of A blocks, the distribution function $f_{\mathrm{A}}(\ell)$ also decays exponentially in the initial region, which corresponds to short blocks included in the random chain sections. For longer A blocks, however, the distribution becomes significantly broader and has a local maximum at $\ell \sim 10$ [95]. Hence, one can conclude that the distribution of A blocks strongly deviates from that known for random sequences.

By varying the distance between nearest adsorption sites, r_{s}, one can control the composition variation period of the synthesized copolymer. From the chemical correlators defined by Eq. 16, it is easy to find the average number of segments in the repeating chain sections, N, for different r_{s} values. It is instructive to analyze the relation between N and r_{s}. As expected, a power law $N \propto r_{\mathrm{s}}^{\mu}$ is observed. It is clear that exponent μ in this dependence should be between $\mu = 1$ (for a completely stretched chain) and $\mu = \nu^{-1}$ with $\nu \approx 0.6$ (for a random coil with excluded volume [75]). The calculation [95] yields $\mu \approx 1.33$ for $N \geq 15$. This supports the aforementioned assumption that the repeating chain sections are strongly stretched between the adsorption sites. The same conclusion can be drawn from the visual analysis of typical snapshots similar to that presented in Fig. 22.

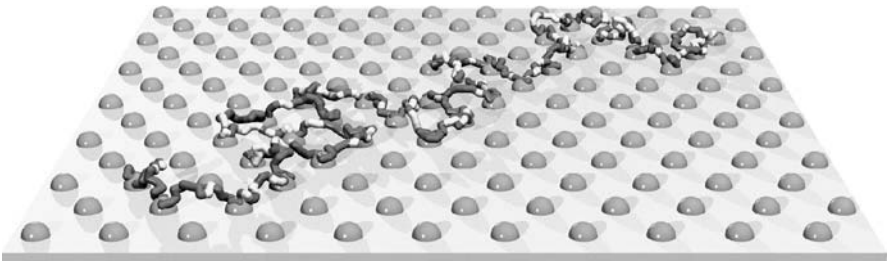

Fig. 22 Snapshot illustrating a typical conformation of 512-unit copolymer chain synthesized near a patterned surface in the strong adsorption regime. Chain segments of A and B type are shown as *dark gray* and *light gray sticks*, respectively. Spheres depict adsorption sites

2.4 Design of Monomeric Units

2.4.1 Amphiphilic Polymers

It is commonplace to say that the properties of a copolymer depend not only on its chemical sequence but also on the chemical structure of its monomeric units. Therefore, the second important route in molecular design can be connected with designing monomeric units of the copolymer having a given sequence distribution. One of the promising ways in this direction is to adjust the *amphiphilic* properties of the copolymer chain.

A large number of macromolecules possess a pronounced amphiphilicity in every repeat unit. Typical examples are synthetic polymers like poly(1-vinylimidazole), poly(*N*-isopropylacrylamide), poly(2-ethyl acrylic acid), poly(styrene sulfonate), poly(4-vinylpyridine), methylcellulose, etc. Some of them are shown in Fig. 23. In each repeat unit of such polymers there are hydrophilic (polar) and hydrophobic (nonpolar) atomic groups, which have different affinity to water or other polar solvents. Also, many of the important biopolymers (proteins, polysaccharides, phospholipids) are typical amphiphiles. Moreover, among the synthetic polymers, polyamphiphiles are very close to biological macromolecules in nature and behavior. In principle, they may provide useful analogs of proteins and are important for modeling some fundamental properties and sophisticated functions of biopolymers such as protein folding and enzymatic activity.

Since amphiphilic polymers contain monomeric units having hydrophobic/hydrophilic character, they can exhibit conformational transitions induced by temperature, solvent composition, or pH variation [96]. Because of the presence of the two opposing interactions towards the solvent in which they are immersed, amphiphiles can self-assemble, forming a variety of supramolecular structures. Understanding the physics of self-association

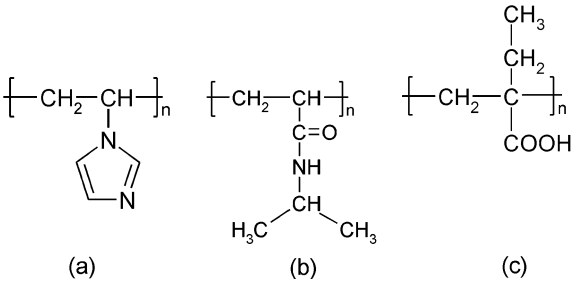

Fig. 23 Examples of amphiphilic polymers: **a** poly(1-vinylimidazole), **b** poly(*N*-isopropyl acrylamide), and **c** poly(2-ethyl acrylic acid)

of amphiphiles is extremely challenging and also important because the underlying ideas have found connections to other fundamental areas, e.g., phase transitions in membranes, crumpled surfaces, and geometry of random surfaces.

2.4.2 HA Model

The two-letter ("black-and-white") HP model first introduced by Lau and Dill [31] and widely discussed in this article is the simplest model of hydrophobic/hydrophilic polymers. The model is very computationally efficient, but its principal disadvantage is the representation of each monomeric unit of an amphiphilic chain as a point-like interaction site of pure hydrophilic or pure hydrophobic type. At the same time, in the large majority of real amphiphilic polymers, *each* monomeric unit has a *dualistic* (hydrophobic/hydrophilic) character, that is, repeating polymer unit, which is considered as hydrophilic, actually incorporates both hydrophilic and hydrophobic parts concurrently. A typical example is poly(1-vinylimidazole) with the hydrocarbon (hydrophobic) backbone and hydrophilic water-soluble side groups (see Fig. 23). Many of the 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 site-site interactions, as it is presupposed in the HP model by discarding all details of side-group interactions.

One of the possible extensions of the HP model is the HA *side chain model* introduced in [212]. This is a more realistic coarse-grained model of amphiphilic polymers where the dualistic character of each monomeric unit is explicitly represented.

In terms of graph theory, an amphiphilic polymer can be modeled as a caterpillar graph instead of a linear graph corresponding to the standard HP model. More formally, a caterpillar of length N is the $\boldsymbol{HP}$ graph in which the set $\{\boldsymbol{H}\}$ represents the nodes in the backbone and the set $\{\boldsymbol{P}\}$ the so-called legs. Each backbone node corresponds to a hydrophobic group (e.g., $CH_2 - CH$ group in Fig. 23) whereas the leg is considered as a polar side group attached to the node. With this representation, each amphiphilic monomer unit (A) is treated as a hydrophobic/hydrophilic HP dumbbell.

Depending on the content of pure hydrophobic and amphiphilic groups, one can simulate amphiphilic homopolymers (poly-A) and copolymers with the same HA composition but with different distribution of H and A units along the hydrophobic backbone, including regular copolymers comprising H and A units in alternating sequence, $(HA)_x$, regular multiblock copolymers $(H_L A_L)_x$ composed of H and A blocks of equal lengths L, and random copolymers having different H and A block lengths (Fig. 24). Gener-

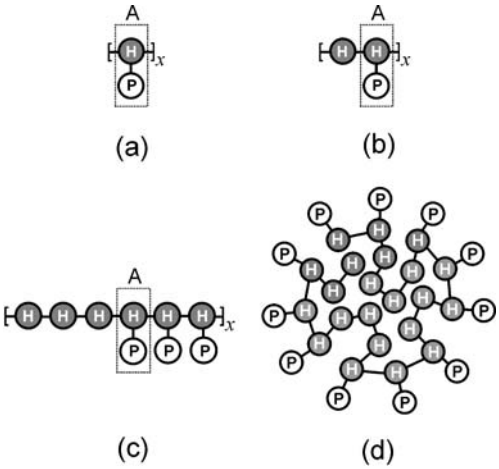

Fig. 24 The "side-chain" models of amphiphilic polymers: **a** amphiphilic homopolymer (poly-A), **b** regular alternating HA copolymer, **c** regular multiblock HA copolymer, and **d** protein-like HA copolymer. Each hydrophobic monomer unit (H) is considered as a single interaction site (bead); each amphiphilic group (A) is modeled by a "dumbbell" consisting of hydrophobic (H) and hydrophilic (P) beads

ally, a random (quasirandom) amphiphilic copolymer is 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.

As we will see in Sect. 3.4, such a relatively trivial modification of the standard HP model can lead to some nontrivial consequences when studying the collapse for the single-chain amphiphilic polymers and their aggregation in solution.

There are many further issues that can be addressed by the model of the kind described here. Clearly, the HA model is amenable to a number of generalizations that allow one to study more sophisticated features of amphiphilic copolymers, including, for instance, backbone stiffness, orientational degrees of freedom, or additional structural constraints such as the saturation of monomer-monomer interactions [98], which are crucial, e.g., for the folding of RNA. Also, it is easy to introduce dipole moments for side H – P bonds and specific directional interactions (like hydrogen bonds) for some of the chain units. These additional factors can result in the formation of intramolecular secondary structures and lead to an increase in the stability of globules formed by such polymers.

In the literature, there are several other coarse-grained polymer models in which spherical monomers are replaced by asymmetric objects. Generally, this gets a host of qualitatively new structures, e.g., liquid crystalline phases of helical secondary structures [99]. In those models, including that

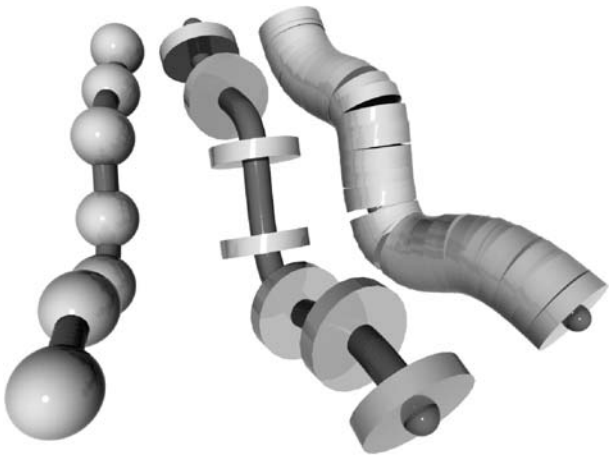

Fig. 25 Chain of beads, chain of coins, and chain of stacked coins

described here, there are two main ingredients: chain units are connected to each other in a linear fashion and each unit of the chain possesses a spatial direction representing the local direction associated with the chain. Familiar examples in protein science include the exotic models that treat the protein backbone not as a chain of spheres but as a chain of anisotropic objects (such as coins) for which one of the three directions differs from the other two (Fig. 25). If such a chain is viewed as being made up of stacked coins instead of tethered spheres, we naturally arrive at the picture of an elastic tube (like a garden hose or spaghetti) whose axis coincides with the chain backbone (Fig. 25). At this coarse-grained level of description, new physics arises from the interplay between two length scales: the range of (anisotropic and many-body) attractive interactions and the thickness of the tube [99].

3
Properties of Designed Copolymers

3.1
Single Chains

3.1.1
Coil-to-Globule Transition

The coil-to-globule transition was studied for designed HP copolymer chains both by means of lattice Monte Carlo simulations using bond fluctuation algorithms and multiple histogram reweighting [100, 101] and by a numer-

ical self-consistent-field (SCF) method [101], assuming that the dominant driving force for polymer collapse is H – H attraction. Copolymer chains of fixed length with H and P monomeric units with regular, random, and specially designed (protein-like) sequences obtained via the coloring procedure were investigated. Qualitatively, the results from both methods are in agreement. The calculations show that it is possible to distinguish four temperature regions.

There exists a low temperature below which the energy of the strongly collapsed chains is constant so that the chains appear to be at the lowest energy. This state may be called the *ground state* or the "frozen glass" phase. Of course, for random copolymers, this state is degenerate, i.e., it is not unique. Its energy is averaged over a few low-energy conformations for each sequence. The type and number of segment-segment contacts in the ground state is well defined for a particular copolymer sequence. With increasing block size L the ground state energy decreases. The reason for this is that small blocks along the chain cannot avoid internal unfavorable H – P contacts. The ground state level of random sequences is significantly higher than that for a protein-like chain. This is because in the protein-like sequence there are longer stretches of hydrophobic units that can efficiently pack in the interior of the dense globule than in the random sequences. Here, we thus note once more that the coloring design procedure [18] results in chains that significantly differ from their random analogues.

At a slightly higher temperature, fluctuations in chain conformations become possible and the compact molecule undergoes internal rearrangements. Under these conditions, the chains remain rather compact but the interaction energy per segment increases more or less linearly with the temperature T. We refer to this regime as the *molten globule state* [102]. There is not a strong L dependence for the onset of the molten globule regime. When more entropic restrictions are locked into the ground state (i.e., for smaller L), the system enters the molten globule state easier. The longer the block length the larger the temperature range over which the molten globules state is found. Both the protein-like and the random copolymers have a well-developed molten globule regime. For protein-like chains this temperature region is wider than the corresponding region for uncorrelated random copolymer chains, because the latter have, on average, shorter stretches of hydrophobic chain segments. The intermediate molten globule state appears most pronounced for the protein-like sequences. Since protein-like sequences behave (quasi)randomly, the calculations [101] thus confirm the previous findings [18–20] that the protein-like copolymers can "inherit" some information from the parent (globular) conformation used to generate its sequence.

With a further increase in temperature, at the end of the molten globule regime, there is a sudden jump in the interaction energy. At this temperature, the coil-to-globule transition is found. The transition temperature increases

with increasing block length. For the regular multiblock copolymers the energy jump appears especially large for chains with intermediate block lengths, and the transition looks like a first-order phase transition. Significantly, it was found that the protein-like and the random copolymers have a very small energy jump at the coil-to-globule transition [101]. The transition thus tends to become continuous with increasing levels of heterogeneity along the chain.

Thus, the cooperativity of the coil-to-globule transition for a copolymer with a given block length is highest for regular block copolymer chains. This is explained by renormalization of effective monomeric unit in the case of monodisperse blocks. The cooperativity is lower for the random and protein-like sequences because of the polydispersity of the block length. For regular multiblock copolymers consisting of alternating H and P blocks, the transition temperature is roughly an exponential function of the block length. It should be emphasized that the transition temperature observed for protein-like sequences is much higher than could be expected from their number average block length. Indeed, these transition temperatures are close to those of regular copolymers in which the block length is several folds larger [101]. Because the correlations in the sequence distribution of protein-like copolymers obey the Lévy-flight statistics [35], very long blocks can occur in their chain if the total chain length allows this to happen. Such long blocks increase the sharpness of the transition.

The globular state for protein-like sequences is found to be more stable than that for statistical random copolymers [101]. As has been noted above, the globules of protein-like copolymers exhibit a dense micelle-like core of hydrophobic H units stabilized by the long dangling loops of hydrophilic P units (see Fig. 3a). From the analysis of the collapse transition, one can also conclude that the protein-like sequence is reflected in the shift of the coil-to-globule transition temperature to higher temperatures as compared to the random-block counterparts with the same composition and that same degree of blockiness [18]. Thus, the numerical studies [18, 100, 101] show that specific sequence correlations play an important role for the chain folding and the stability of heteropolymer globules.

Finally, at very high temperature the chains are swollen and behave as a random coil. In such a chain the majority of the interactions are between polymer units and solvent molecules. This state does not depend on the primary structure of the chains and, therefore, the interaction energy levels off to a constant value that depends only on chemical composition.

The effect of copolymer sequence on coil-to-globule transition was also studied using Langevin molecular dynamics [103]. The method for estimation of the quality of reconstruction of core-shell globular structure after chain collapse was proposed. It was found that protein-like sequences exhibit better reconstruction of initial globular structure after the cooling procedure, as compared to purely random sequences.

3.1.2 Kinetics of the Collapse Transition

Experimental observation of the coil-to-globule transition in polymers is very difficult because at concentrations accessible to experiments, the intramolecular collapse of single chains competes with the intermolecular aggregation of macromolecules [104–111]. Consequently, there have been only a few studies on the kinetics of the collapse transition. One possible experimental technique to follow the polymer collapse is to measure the time evolution of the gyration radius by light scattering of a highly dilute solution (in order to get single molecule properties). In computer experiments, we do almost the same observations.

For proteins, the problem is seen as follows: "When an egg is boiled, the proteins it contains unfold. Can this procedure be reversed in theory? Or, in other words, can the encrypted code of protein folding be deciphered from the sequence?" [112, 113].

The "fastest" proteins fold amazingly quickly: some as fast as a millionth of a second. While this time is very fast on a person's timescale, it is remarkably long for computers to simulate. In fact, there is about a 1000-fold gap between the simulation timescales and the times at which the fastest proteins fold. This is why the simulation of collapse kinetics is extremely computationally demanding. Thus, the current challenge lies in understanding how particular chemical sequences in coarse-grained copolymer models lead to particular collapse features. This is a fundamental issue in the problem.

The formation of a compact globule in copolymers requires them to have specific conformations, which are reached through local conformational fluctuations. A characteristic collapse time may be defined as for instance the time for which the gyration radius reaches its equilibrium value. This time measures the approach to equilibrium for the system and is related to the mean first passage time.

The main picture that has emerged from theory and different numerical studies is that polymer collapse proceeds as a series of steps, first the rapid formation of small clusters ("blobs") distributed along the chain that grow and then merge, then the formation of a globular state [114–120]. Of course, the features of the collapse transition depend on the quench depth and the chain rigidity. For heteropolymers, we should also take into account their sequence and chemical composition. Numerical research has attempted to design copolymer sequences with desired properties (e.g., rapid folding to a compact globule) in a manner that mimics biological evolution [26, 121]. An interesting feature of these designed sequences is that they often display a well-defined structure in which similar monomers associate in blocks. Recently, it has been shown that intrachain microphase separation plays an important role in both the thermodynamics and kinetics of collapse for HP heteropolymers with various block sizes [122].

Usually, the simulations of folding kinetics involve an initial equilibration period at sufficiently high temperatures, starting from an initial chain conformation in the strongly swollen state, followed by an instantaneous quench to low temperatures (or by an abrupt decrease in the solvent quality) at which the polymer collapses towards a final state. The collapse is monitored using, e.g., the time-dependent radius of gyration $R_g(t)$ as a function of time measured from the moment of the quench. The degree of compactness of the chain can be monitored by examining how the $R_g(t)$ value changes as folding progresses. It is convenient to use the normalized quantities, $Q^* = Q(t)/Q(0)$, that determine the departure of a given time-dependent property $Q(t)$ from its value in the initial state. To gain better statistics, many independent collapses ($\sim 10^2$) are simulated for each copolymer sequence, each beginning with a unique chain conformation. With a uniform probability density distribution in the initial state, the folding time, t^*, is the mean first passage time (MFPT) to the final compact state, which is defined to be the average over many independent simulations of the time to achieve the final state. Also, one can define the mean intermediate folding times, e.g., $\langle t_{1/4} \rangle$, $\langle t_{1/2} \rangle$, etc., which are required to achieve the corresponding intermediate values $Q^*/4$, $Q^*/2$, etc. For $R_g(t)$, $\langle t_{1/2} \rangle$ is the mean time required to halve the chain size.

Let us compare the kinetics of the selective-solvent-induced collapse of protein-like copolymers with the collapse of random and random-block copolymers [18]. Several kinetic criteria were examined using Langevin molecular dynamics simulations. There are some general results, which seem to be independent of the nature of interactions or the kinetic criteria monitored during the collapse. Here, we restrict our analysis to the evolution of the characteristic ratio $\zeta = \left(R^2_{gP}/R^2_{gH}\right)^{1/2}$ that combines the partial mean-square radii of gyration calculated separately for hydrophobic and hydrophilic beads, R^2_{gH} and R^2_{gP}. This ratio takes into account both the properties of compactness and solubility for a heteropolymer globule [70] (compactness is directly related to the mean size of the hydrophobic core, whereas solubility should be dependent on the size of the hydrophilic shell).

Using the ζ value, let us define the time-dependent quality $2\,[\zeta(t)/\zeta(0) - 1]$ and three intermediate folding times $t_{1/4}$, $t_{1/2}$, and $t_{3/4}$, describing its evolution, as well as the corresponding sequence-averaged probability distribution functions $W_{1/4}$, $W_{1/2}$, and $W_{3/4}$. The distribution of folding times averaged over 1000 different sequences of 128-unit HP copolymers with random, random-block, and protein-like statistics are shown in Fig. 26.

It is seen that among different protein-like sequences, we can find significantly larger fractions of the sequences having small characteristic collapse times than for random and random-block sequences with the same 1 : 1 HP composition. This is true both in the early stages of collapse and in the late stages. Those sequences can be called “fast folders”. In a sense, dividing the

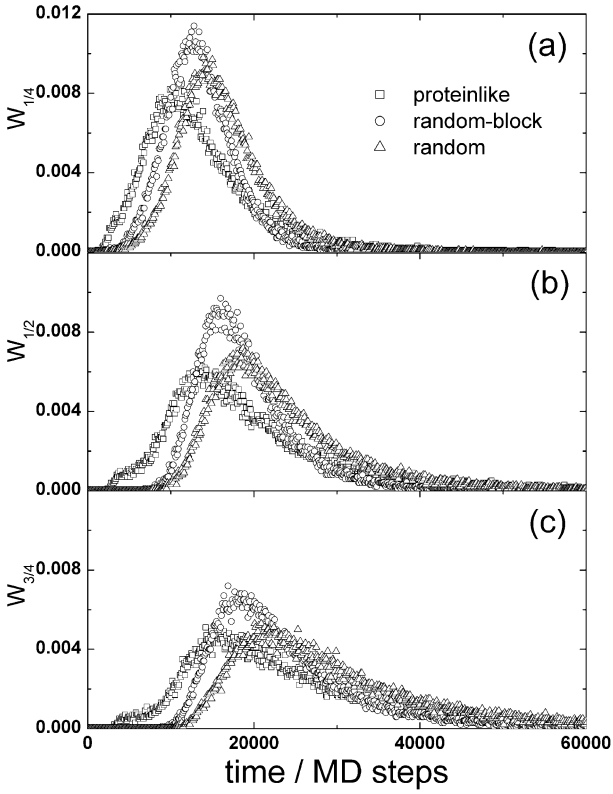

Fig. 26 Distributions of folding times for 128-unit HP copolymers with protein-like, random-block, and random statistics

sequences by fast folders and "slow folders" (that is, those laying before and after the maximum of the W function, respectively) corresponds to a kinetic sequence design or to choosing from the spectrum all possible sequences a set whose kinetic properties possess the attributes we desire. From this viewpoint, our computational procedure is a specific sieve that separates "good" sequences and "bad" ones.

Therefore, we conclude that for the designed copolymers, collapse occurs at a markedly higher kinetic rate, $k \sim 1/\langle t_{1/2} \rangle$. The kinetic properties are clearly related to the sequence distribution. Still it is possible to select "champions" from the protein-like fast folders. Who are those champions? They normally have a larger mean length of hydrophobic and hydrophilic blocks, wider distribution of block length and the average hydrophobicity decreasing towards the chain ends.

Of course, many random processes generating long sequences might in principle produce fast folders, however, there is a negligible chance that we would be able to find their noticeable fraction.

3.2 Phase Behavior: The Sequence-Assembly Problem

Heteropolymers can self-assemble into highly ordered patterns of microstructures, both in solution and in bulk. This subject has been reviewed extensively [1, 123–127]. The driving force for structure formation in such systems is competing interactions, i.e., the attraction between one of the monomer species and the repulsion between the others, on the one hand, and covalent bonding of units within the same macromolecule, on the other hand. The latter factor prevents the separation of the system into homogeneous macroscopic phases, which can, under specific conditions, stabilize some types of microdomain structures. Usually, such a phenomenon is treated as microphase separation transition, MIST, or order-disorder transition, ODT.

For diblock copolymers, periodically arranged spheres (micelles), hexagonally packed cylinders, and a lamellar phase have been observed [1]. A more complex bicontinuous cubic phase with $QI_{a\bar{3}d}$ symmetry (gyroid structure) has also been identified. These supramolecular structures, with length scales on the order of 1 to 10^2 nm, may be controlled by changing the amount of solvent, the length of blocks, or the proportions of A and B monomeric units [128–131].

The conventional theoretical approach to determine the three-dimensional organization of polymers at equilibrium is based on minimizing a suitable Ginzburg–Landau free energy functional with respect to the corresponding lattice parameters for some *a priori* presupposed structure of ideal symmetry (e.g., body centered cubical, hexagonal, lamellar, etc.) [1, 132, 133]. Because of its inherent assumed symmetries, this approach fails to predict the irregular and "defect" microstructures. Also, the computations of the free-energy coefficients within the Leibler-like or random phase approximation (RPA) treatment [132] do not take into account directly the interaction potentials of polymer segments and the system-specific features of polymer architecture on short length scales. Self-consistent mean-field (SCMF) techniques have been proposed in the context of field-theoretic ideas by Matsen and coworkers [134–137] and Fredrickson and coworkers [138, 139]. Microscopic statistical mechanical polymer theories have also been developed, including the polymer RISM theory [140–143], which is an extension to flexible long-chain polymers of the integral equation reference interaction site model (RISM) theory of Chandler and Andersen [144], and density functional (DF) theory [145].

The theoretical approaches mentioned above mostly deal with diblock or regular multiblock AB copolymers [1, 123–125]. It has been shown theoretically [146–149] that for random copolymers the existence of ordered microphases is also not excluded. In particular, it has been shown that in addition to the classical microdomain superstructures, chemical irregularities in multiblock copolymers may result in formation of very exotic positionally ordered structures with many levels (length-scales) of organization

(the so-called "secondary superstructures") [150–152]. The phase behavior of random AB copolymers with specific statistics and a strongly correlated distribution of units along the chain is of particular interest. Treating processes taking place in these systems continues to challenge theorists. Even ideal random copolymer mixtures, produced if the reactivities of the two polymerizing comonomers are equal, have unique statistics: each sequence represents a different component. In general case, the possible random sequences generate mixtures of $2^{N-1} + 2$ chemically different N-unit macromolecules.

In this section, we discuss the phase behavior of protein-like copolymers that can be considered as a specific type of correlated copolymers. One can expect that the presence of long-range correlations in protein-like chains will influence the phase behavior of such copolymers.

The main parameters characterizing the thermodynamic states (segregation regimes) of self-assembling AB copolymers are the composition, $\varphi_i = N_i/N$ (i = A, B; $N_A + N_B = N$; N being the total chain length) and the Flory–Huggins parameter χ, which is inversely proportional to temperature, which reflects the interaction between different segments. For random copolymers, we should also introduce an additional parameter, viz., the average length of the segments composed of units A and B, L_A and L_B. We restrict our consideration to the simplest symmetric case: $\varphi_A = \varphi_B = 1/2$ and $L_A = L_B = L$. For comparison, it is instructive to consider a random-block copolymer with the same composition in which the distribution of block length ℓ is described by the Poisson law. In this case, L can vary. For both copolymers, the length of the segment composed by the units of a given type is a random quantity. It is of interest to compare the behavior of such polymers with each other and with regular multiblock copolymers where L is fixed.

3.2.1 The Polymer RISM Theory

To describe the equilibrium structure of the system, one can use the polymer integral equation RISM theory [140, 141], which allows one to find collective correlation functions. For AB copolymers, the polymer RISM equation is represented in the matrix form [142, 143]

$$\boldsymbol{H}(r) = \int_{(\boldsymbol{r}')} \int_{(\boldsymbol{r}'')} \boldsymbol{W}(|\boldsymbol{r} - \boldsymbol{r}'|)\boldsymbol{C}(|\boldsymbol{r}' - \boldsymbol{r}''|)[\boldsymbol{W}(\boldsymbol{r}'') + \rho\boldsymbol{H}(\boldsymbol{r}'')]\mathrm{d}\boldsymbol{r}'\mathrm{d}\boldsymbol{r}'' . \tag{17}$$

Here, $\boldsymbol{H}$ and $\boldsymbol{C}$ are symmetric matrices whose elements are the partial total $h_{\alpha\beta}(r)$ and direct $c_{\alpha\beta}(r)$ pair correlation functions (α,β = A,B); $\boldsymbol{W}$ is the matrix of intramolecular correlation functions $w_{\alpha\beta}(r)$ that characterize the conformation of a macromolecule and its sequence distribution; and ρ is the average number density of units in the system. Equation 17 is complemented by the closure relation corresponding to the so-called molecular Percus–

Yevick approximation [141]

$$\int_{(\boldsymbol{r}')}\int_{(\boldsymbol{r}'')} \boldsymbol{W}(|\boldsymbol{r}-\boldsymbol{r}'|)\boldsymbol{C}(|\boldsymbol{r}-\boldsymbol{r}''|)\boldsymbol{W}(\boldsymbol{r}'')\mathrm{d}\boldsymbol{r}'\mathrm{d}\boldsymbol{r}'' \tag{18}$$

$$= \int_{(\boldsymbol{r}')}\int_{(\boldsymbol{r}'')} \boldsymbol{W}(|\boldsymbol{r}-\boldsymbol{r}'|)[\boldsymbol{C}^{(0)}(|\boldsymbol{r}'-\boldsymbol{r}''|) + \Delta\boldsymbol{C}(|\boldsymbol{r}'-\boldsymbol{r}''|)]\boldsymbol{W}(\boldsymbol{r}'')\mathrm{d}\boldsymbol{r}'\mathrm{d}\boldsymbol{r}'',$$

$$r > \sqrt{\sigma_\alpha\sigma_\beta}$$

$$\Delta c_{\alpha\beta}(r) = [1 - \mathrm{e}^{-u_{\alpha\beta}(r)/kT}][h^{(0)}_{\alpha\beta}(r) - 1], \quad r > \sqrt{\sigma_\alpha\sigma_\beta} \tag{19}$$

$$h_{\alpha\beta}(r) = -1, \quad r < \sqrt{\sigma_\alpha\sigma_\beta} \tag{20}$$

$$c^{(0)}_{\alpha\beta}(r) = 0, \quad r > \sqrt{\sigma_\alpha\sigma_\beta} \tag{21}$$

Here, $u_{\alpha\beta}(r)$ describes the interaction between nonbonded units and σ_α is the effective unit size ($\sigma_\mathrm{A} = \sigma_\mathrm{B} = \sigma$). In [153, 154], the Yukawa potential was used

$$u_{\alpha\beta}(r) = \begin{cases} \varepsilon_{\alpha\beta}(\sigma/r)\exp[-2(r/\sigma - 1)], & r \geq \sigma \\ \infty, & r < \sigma \end{cases}, \tag{22}$$

where $\varepsilon_{\alpha\beta}$ is an energy parameter which is directly related to the χ parameter. We consider the case when A units are attracted to one another ($\varepsilon_{\mathrm{AA}} = -1$) while B-B and A-B interactions are the excluded-volume type ($\varepsilon_{\mathrm{AA}} = \varepsilon_{\mathrm{AA}} = 0$). The functions $h^{(0)}_{\alpha\beta}(r)$ and $c^{(0)}_{\alpha\beta}(r)$ in Eqs. 18–21 correspond to the reference (athermal) system for which all $\varepsilon_{\alpha\beta} = 0$. Within the frame of Gaussian chain statistics, the spatial distribution of the pairs of units i and j separated by $n = |i - j|$ bonds in a given chain is written in the reciprocal $\boldsymbol{q}$-space as $w_{ij}(q) = [\sin(q\sigma)/q\sigma]^n$. Summation of $w_{ij}(q)$ over i and j ($i,j = 1, 2, ..., N$) gives the matrix elements $w_{\alpha\beta}(q)$ usually termed single-chain form-factors. Below, the σ and $\varepsilon_{\mathrm{AA}}$ values are used as basic units.

The polymer RISM theory provides the local structural information through the site-site pair correlation functions (PCFs) and the influence of the long wavelength structure for these local properties. Although the phase-separated structures are not available from the polymer RISM theory, one can obtain the structural information for the disordered single-phase at a certain distance from the phase separation point against the athermal reference system. Also, by varying temperature T and monomer density ρ, one can find the conditions under which the spatially homogeneous state of the system becomes unstable. This takes place on a spinodal line, which is determined by the set $\{T^*,\rho^*\}$ or parametrically as $T^* = T^*(\rho)$. The condition of spinodal instability at a temperature T^* corresponds to [141]

$$\Delta(q) \equiv \det[\boldsymbol{E} - \rho\boldsymbol{W}(q)\boldsymbol{C}(q)]_{q=q^*} = 0, \tag{23}$$

where $\Delta(q)$ is the determinant of the matrix equation (Eq. 17), q^* represents the wave vector of maximum instability, and $\boldsymbol{E}$ is the unit diagonal matrix.

The value $q = q^*$ at which $\Delta(q^*) \to 0$ allows one to estimate the spatial scale r^* of the arising structure ($r^* = 2\pi/q^*$).

Figure 27 shows the plots of T^* and r^* versus L for the regular and random-block copolymers at the volume fraction of macromolecules $\Phi = 0.8$ corresponding to a typical polymer melt. It is seen that T^* increases in both cases with an increase in the block length; however, at small L values, the T^* values for the random-block copolymer are higher than for the regular one. In other words, structure formation in the system of copolymers with a random distribution of block lengths occurs more readily (i.e., at a higher temperature) than in the system of regular multiblock copolymers. This result, which might appear at first glance to be surprising, will be interpreted in Sect. 3.2.3.

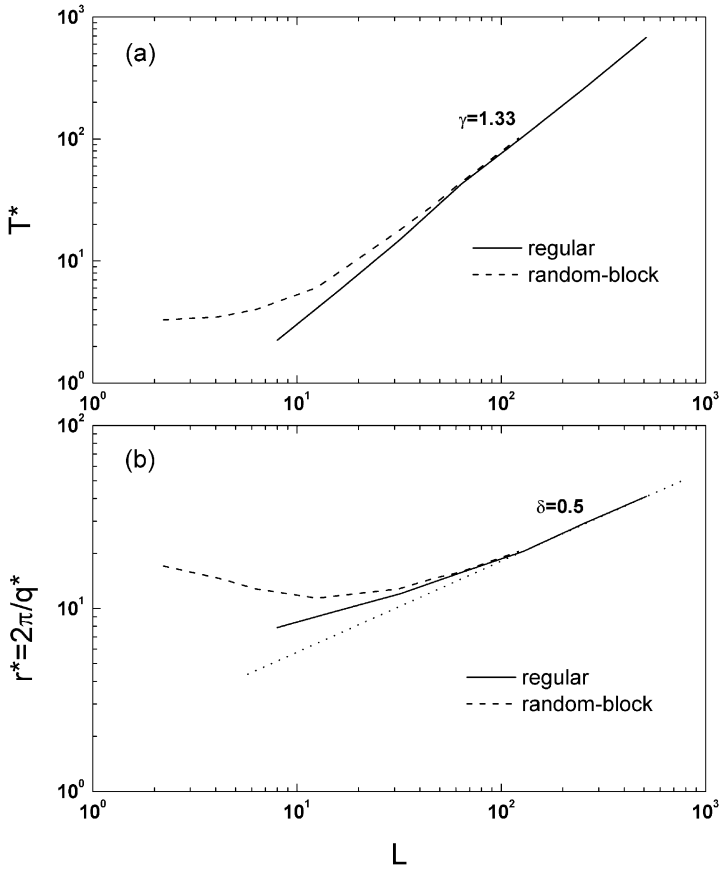

Fig. 27 The values of T^* and r^* as a function of block length L for the regular and random-block copolymers at the volume fraction of macromolecules $\Phi = 0.8$. Adapted from [153]

When the block length becomes comparable with N, distinctions between the behaviors of two copolymers practically disappear. At $L > 200$, one observes that $T^* \sim L^{\gamma}$ with $\gamma = 4/3$. In this case, the characteristic scale of the microdomain structure behaves as $r^* \sim L^{\delta}$ with $\delta = 1/2$. This dependence is caused by the fact that flexible chains in the melt have a Gaussian conformation, and the average size of any chain section of n units is proportional to $n^{1/2}$ [75]. Hence, for sufficiently large L's, the spatial scale of microinhomogeneities in the system is determined only by the block size. However, the behavior of the random-block copolymer at $L \lesssim 10^2$ is more complicated. In particular, r^* has a minimum at $L \approx 10$.

Figure 28 shows the spinodal lines for all the systems under discussion. At a low Φ ($\lesssim 0.1$), no noticeable differences are observed. This region of the phase diagram (including the critical point) corresponds to macrophase sep-

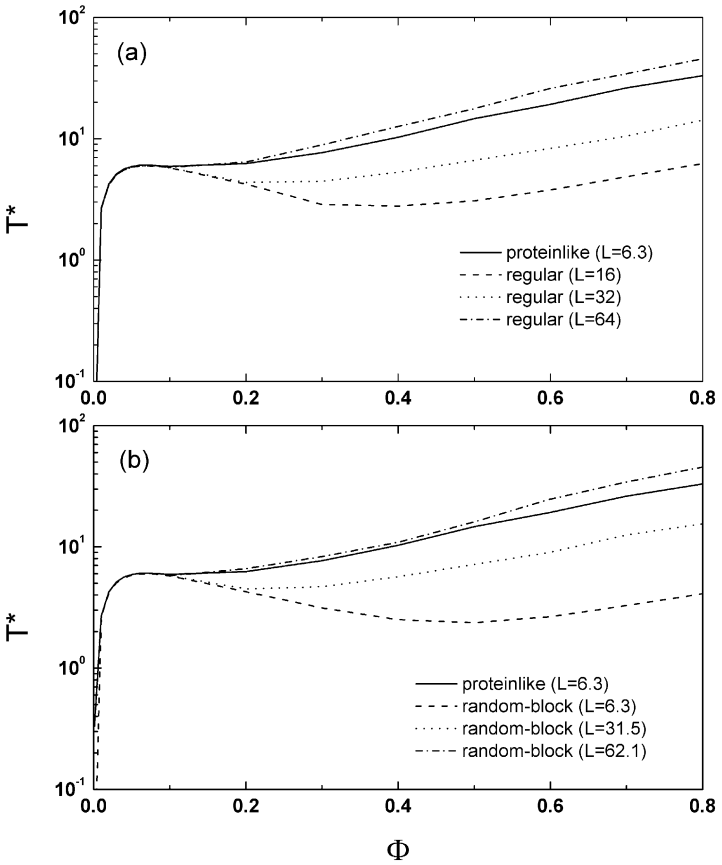

Fig. 28 Spinodals for the systems of protein-like copolymers with average block length L = 6.3 and of regular and random-block copolymers having different block lengths. Adapted from [153]

aration transition, MAST, i.e., to the conditions when determinant (Eq. 23) vanishes at $q^* = 0$. In this case, specific features of copolymer sequence have virtually no effect. However, there are such Φ values at which the condition $\Delta(q^*) \to 0$ is met at $q^* \neq 0$ (Fig. 29). Transition from the $q^* = 0$ regime to the $q^* \neq 0$ regime occurs at the Lifshitz point delimiting the MAST and MIST regions.

In the $q^* \neq 0$ region, the phase behavior of protein-like copolymers begins to sharply differ from that observed for the regular and random-block copolymers. In particular, Fig. 28 shows that the spinodal line for the 1024-unit protein-like copolymers with the average block length $L = 6.3$ is close to the spinodals of the regular and random-block copolymers in which the block length is roughly tenfold (!) larger [153]. At the same Ls, the transition temperatures for the protein-like copolymers are also several fold larger. This indicates that the processes of self-organization in the system of protein-like copolymers can proceed significantly more intensely than in the other copolymer systems. We note that a similar behavior is observed for intramolecular (coil-to-globule) transitions discussed above. Moreover, Fig. 29 shows that at a sufficiently high density of protein-like copolymers ($\Phi \gtrsim 0.3$), the domain spacing $r^* = 2\pi/q^*$ noticeably exceeds that observed for the random-block copolymers. The difference between the protein-like and regular copolymers is even greater.

Therefore, the PRISM calculations [153] show that the protein-like copolymers with A-A attractions are more "inclined" to self-organization than their random-block counterparts and regular multiblock copolymers with the same chemical composition and the same chain length.

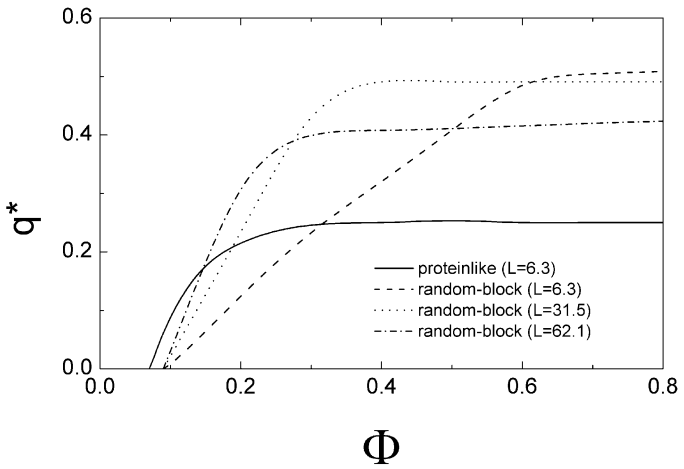

Fig. 29 Characteristic wave number q^* as a function of polymer volume fraction for the systems of protein-like copolymers with $L = 6.3$ and random-block copolymers with different block lengths. The domain spacing is defined as $r^* = 2\pi/q^*$. Adapted from [153]

3.2.2 Field-Theoretic Calculation

The RISM theory describes copolymers only in the weak-segregation limit. At much stronger interaction of monomeric units, the components should be strongly segregated, with a narrow interphase between domains. In this case, the self-consistent field (SCF) methods [134–139] can be used. They span both weak- and strong-segregation limits, as well as the intermediate regime where the other analytical approaches do not apply. Here we present some results obtained by using the real-space SCF approach [155] that is particularly well suited to screening for new types of self-assembly in copolymer melts because it requires no assumption of the mesophase symmetry.

In the SCF theory, the external mean fields acting on a polymer chain are calculated self-consistently with the composition profile. Each chain has independent statistics in average chemical potential fields, $\omega_\alpha(\boldsymbol{r})$, conjugate to the volume fraction fields, $\phi_\alpha(\boldsymbol{r})$, of monomer species α. The free energy per chain F is related to the statistical weight, $q(\boldsymbol{r},\boldsymbol{s})$, that a segment of a chain, originating from the free end of the α block and with contour length $\boldsymbol{s}$, has at its terminus at point $\boldsymbol{r}$. The free energy is to be minimized subject to the constraint of local incompressibility and the constraint that $q(\boldsymbol{r},\boldsymbol{s})$ satisfies a dynamical trajectory given by modified diffusion equations. Minimization of F with respect to $\omega_\alpha(\boldsymbol{r})$ and $\phi_\alpha(\boldsymbol{r})$ leads to a self-consistent set of equations [155] that are solved iteratively. The propagators $q(\boldsymbol{r},\boldsymbol{s})$, together with

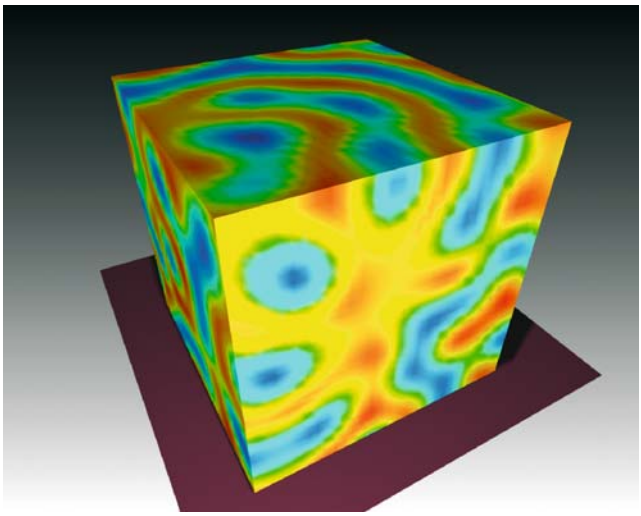

Fig. 30 Composition variations for the incompressible melt of 128-unit protein-like chains at $\chi N = 64$. Regions rich in A and B segments are shown in *red* and *blue*, respectively; intermediate regions are given in *yellow* and *green*

their conjugate propagators $q^{\dagger}(\boldsymbol{r},\boldsymbol{s})$, which propagate from the opposite end of the chain, are used to calculate the density fields. The calculations result in the space distribution of the volume fractions $\phi_{\alpha}(\boldsymbol{r})$.

As an example, we show in Fig. 30 the composition variations that are found for the incompressible melt of 128-unit protein-like chains at $\chi N = 64$, starting from a homogeneous state. The simulation box is a three-dimensional 32^3 grid with periodic boundary conditions and a side length of 6.4 (in units of R_g). Regions rich in A and B segments are shown in red and blue, respectively; intermediate regions are given in yellow and green.

As seen, the SCF theory predicts irregular bicontinuous morphology with no long-ranged order. Note that for random-block copolymers with the same average block length, no microphase separation is observed even for $\chi N = 100$.

3.2.3 Molecular Dynamics Simulation

In molecular dynamics simulations, the mean values are estimated as arithmetic averages over the configurations stored in each trajectory and over all trajectories generated for each system. With this method, thermal fluctuations are taken into account in contrast to the standard RPA model where properties are calculated over static minimum energy configurations. This leads to more reliable estimates of properties at the temperature of interest, including the strong segregation regime.

We will briefly discuss the molecular dynamics results obtained for two systems—protein-like and random-block copolymer melts— described by a Yukawa-type potential with (i) attractive A-A interactions ($\varepsilon_{\mathrm{AA}} < 0$, $\varepsilon_{\mathrm{BB}} = \varepsilon_{AB} = 0$) and with (ii) short-range repulsive interactions between unlike units ($\varepsilon_{AB} > 0$, $\varepsilon_{\mathrm{AA}} = \varepsilon_{\mathrm{BB}} = 0$). The mixtures contain a large number of different components, i.e., different chemical sequences. Each system is in a randomly mixing state at the athermal condition ($\varepsilon_{\alpha\beta} = 0$). As the attractive (repulsive) interactions increase, i.e., the temperature decreases, the systems relax to new equilibrium morphologies.

The partial scattering functions $S_{\mathrm{A}}(q,\varepsilon)$ normalized by $S_{\mathrm{A}}(q,0)$ obtained at the athermal condition are shown in Fig. 31 for both systems at a few selected $\varepsilon_{\mathrm{AA}}$ values. When $\varepsilon_{\mathrm{AA}} \lesssim -1$, the scattering intensity at small, nonzero wave numbers q rises sharply with decreasing $\varepsilon_{\mathrm{AA}}$. Both systems show a common feature in that the peaks in $S_{\mathrm{A}}(q)$ increase and their position q^*, which is directly related to the periodicity of the concentration fluctuations, shifts to lower q values as attraction becomes stronger, indicating that the structure becomes better defined. These results reflect that the mechanisms of the structure development caused by microphase separation are similar to each other. We thus see that both copolymers exhibit intermolecular aggregation as the ordering transition is approached, although we do not find a *periodic* microphase forming in both systems.

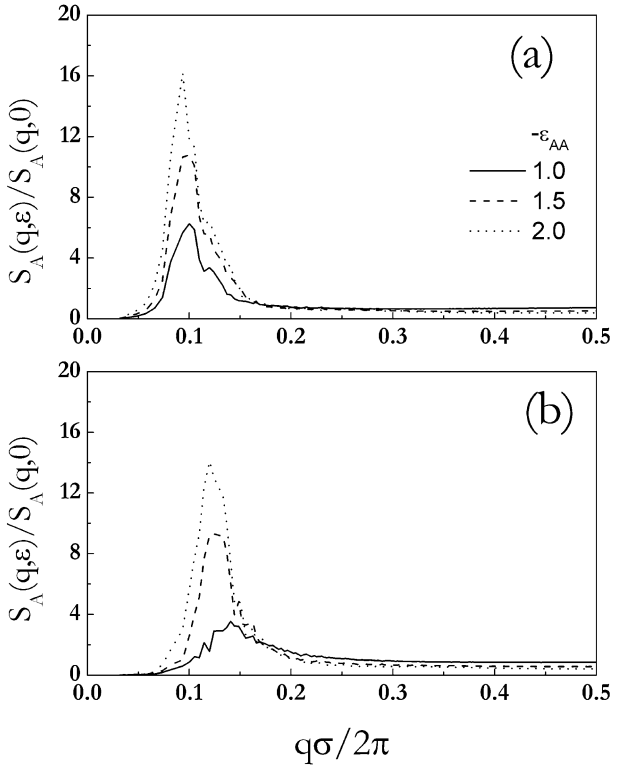

Fig. 31 Normalized partial scattering functions $S_A(q,\varepsilon)$ for the melt-like systems of **a** protein-like copolymers and **b** random-block copolymers at a few selected ε_{AA} values that characterize A-A attraction

To visualize the three-dimensional structures of these microphases more clearly, the density of A-segments was measured on a three-dimensional grid. In the A-rich domains the density of A-segments is high (~ 1), and in the B-rich domains this density is low (~ 0). The dividing surface between the A-rich domains and the B-rich domains is now represented by the isosurface where the density is midway between these values, i.e., 0.5.

When Fig. 32a is compared with Fig. 32b, it is revealed that the domain spacing in the protein-like copolymer system is longer than that in the random-block one. This observation is consistent with the results shown in Fig. 31. The same conclusion can be drawn from the molecular dynamics simulation of the systems with repulsive interactions between unlike units. As an example, we show the normalized scattering functions calculated at large incompatibility, $\varepsilon_{AB} = 1$ (Fig. 33), and the corresponding isosurfaces (Fig. 34). In Figs. 32 and 34, we observe irregular bicontinuous structures with no long-ranged order. Similar morphologies have been predicted in the SCF calculation discussed above as well as in Monte Carlo simulations of random

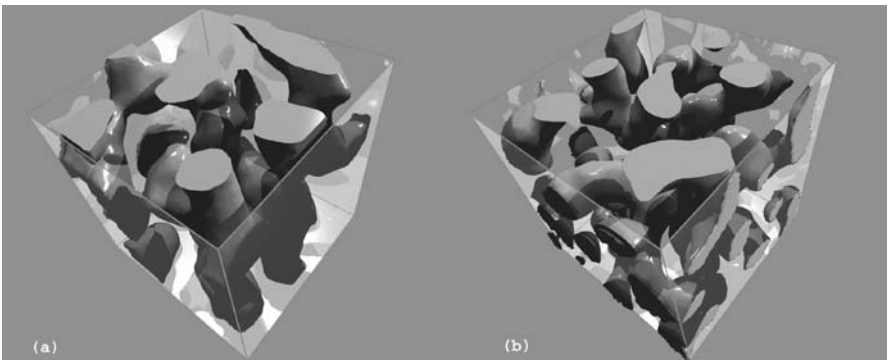

Fig. 32 Isosurfaces showing microphase separated systems of **a** protein-like and **b** random-block copolymers with attractive A-A interactions

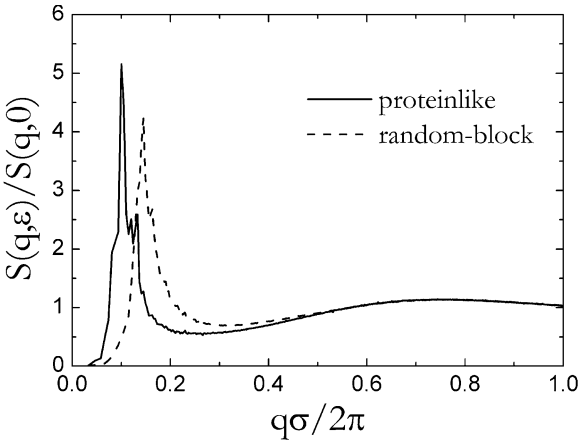

Fig. 33 Normalized scattering functions $S(q,\varepsilon)$ for the melt-like systems of protein-like and random-block copolymers with repulsive A-B interactions

copolymer melts [156, 157]. The reason is that the strong fluctuations destroy completely the stability of the ordered phases, and the disordered phase becomes thermodynamically stable everywhere. In this respect the morphology below the transition temperature resembles a disordered microstructure having an anomalously large correlation length (random wave structure) [158]. This is unique to random copolymer systems. It is clear that this structure should get more and more ordered as the randomness decreases.

What is most important for our discussion is the fact that the spatial scale r^* of the segregated structure for protein-like copolymers is appreciably larger than that for random-block copolymers with the same composition and the same average block length. Also, MIST in the protein-like copolymer system occurs at a temperature higher than that of the random-block system, which is in agreement with the prediction of the polymer RISM theory [153].

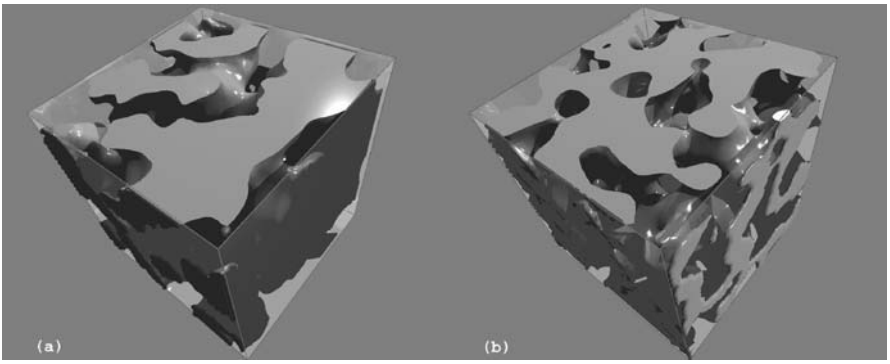

Fig. 34 Isosurfaces showing microphase separated systems of **a** protein-like and **b** random-block copolymers with repulsive A-B interactions

Obviously, these features arise from the difference between the chemical sequences of the two copolymers.

The reason behind this distinction is that the behavior of random copolymers is governed not only by the average block length L but also by the block length dispersion D_L. For correlated random copolymers, the structure formation is dominated by long blocks whose probability increases with an increase in D_L. Even at a relatively low fraction of these blocks in the chain, their effect can be decisive. As we noted while discussing Fig. 27, the largest differences between the regular and random-block copolymers manifest themselves just at relatively small L values when the chain consists of both the short and rather long chemically homogeneous sections. With an increase in L at fixed N, the width of the block length distribution decreases, and the differences gradually disappear (Fig. 27). As has been discussed in Sect. 2.2.2, the block length distribution in protein-like copolymers is described by a specific type of statistics, namely by the Lévy-flight statistics [35] with a high dispersion and a slow decrease in the block length probability. That is why the chains contain a significant fraction of long sections composed of chemically identical segments even at relatively small L values. This is precisely the major reason behind the rather unusual behavior of such copolymers in self-organization processes.

3.2.4 Evolutionary Approach

Shakhnovich and Gutin [26] (see also [28, 159, 160]) showed that it is possible to design a copolymer in such a way that it will fold into a specific conformation. To do this, they optimized the sequence, using a Monte Carlo method that randomly exchanges monomers within the sequence. In this section, we describe an extension of this approach to design a copolymer sequence that is

capable of forming microphase-separated structures with the domain spacing r^* which is as large as possible.

To analyze the stability of the ordered microphases, the simplest incompressible random-phase approximation [132] can be employed. Using this approach, the critical value of the Flory–Huggins parameter, χ^*, and the corresponding spinodal temperature, $T^* = 1/\chi^*$, can be determined by the condition that the scattering intensity $S(q)$ reaches its maximum value at a nonzero wave vector q^*. Within the RPA the scattering intensity is given by [132, 142]

$$S_{\mathrm{RPA}}^{-1}(q) = \mathfrak{R}(q) - 2\chi \tag{24}$$

with

$$\mathfrak{R}(q) = \frac{1}{\delta w(q)} \left\{ \frac{w_{\mathrm{AA}}(q)}{\phi_{\mathrm{B}}} + \frac{w_{\mathrm{BB}}(q)}{\phi_{\mathrm{A}}} + \frac{2w_{\mathrm{AB}}(q)}{f_{\mathrm{A}} f_{\mathrm{B}}} \right\}, \tag{25}$$

where $f_{\mathrm{A}} = N_{\mathrm{A}}/N$, $f_{\mathrm{B}} = N_{\mathrm{B}}/N$, ϕ_{A} and ϕ_{B} are the volume fractions of the corresponding species, and $\delta w(q)$ is defined as $\delta w(q) = w_{\mathrm{AA}}(q) w_{\mathrm{BB}}(q) - f_{\mathrm{A}}^{-1} f_{\mathrm{B}}^{-1} w_{\mathrm{AB}}^2(q)$. To a first approximation, one can consider macromolecules on the basis of the highly simplified unperturbed model without intramolecular excluded volume interactions. This allows us to considerably simplify the problem by calculating the matrix elements $w_{\alpha\beta}(q)$ using the Gaussian intramolecular correlation function $w_{ij}(q) = \exp\left(-q^2\sigma^2 \left|i-j\right| /6\right)$, which characterizes the distribution of segments i and j belonging to the species A and B inside a polymer.

In the model [161], Monte Carlo simulations were performed to search for point mutations that favor the microphase separation of copolymer melt. A random AB sequence is taken as an initial state, and then a procedure of the evolution (annealing) of the sequence starts. The iterative procedure consists of many mutation steps. At each Monte Carlo step, two monomers are chosen randomly and, if they happen to be of different types, an attempt is made to exchange their types (A $\leftrightarrow$ B). This changes the copolymer sequence and the matrix elements $w_{\alpha\beta}(q)$. The resulting change in the transition temperature ΔT^* after an attempted mutation is calculated and the probability p to fix the mutation is guided by the Metropolis algorithm: p equals 1 if $\Delta T^* \geq 0$, otherwise p equals $\exp(\Delta T^*/T_{\mathrm{s}})$, where T_{s} is the fictitious temperature referred to as "sequence design temperature" that characterizes the tolerance to mutations in sequence space. In fact, this is the same parameter as that used in Sect. 2.2.5.

Figure 35 shows the transition temperature T^* and the domain size r^* ($= 2\pi/q^*$) as a function of T_{s} for different N. Note that both T^* and r^* are averaged over the ensemble of generated sequences ($\sim 10^6$). If T_{s} is too high, the sequences tend to become random. In contrast, when T_{s} is too low, the evolutionary algorithm leads to the trivial diblock sequence. In this case, one observes a typical mean-field behavior: $T^* \propto N$ and $r^* \propto N^{1/2}$ (for long

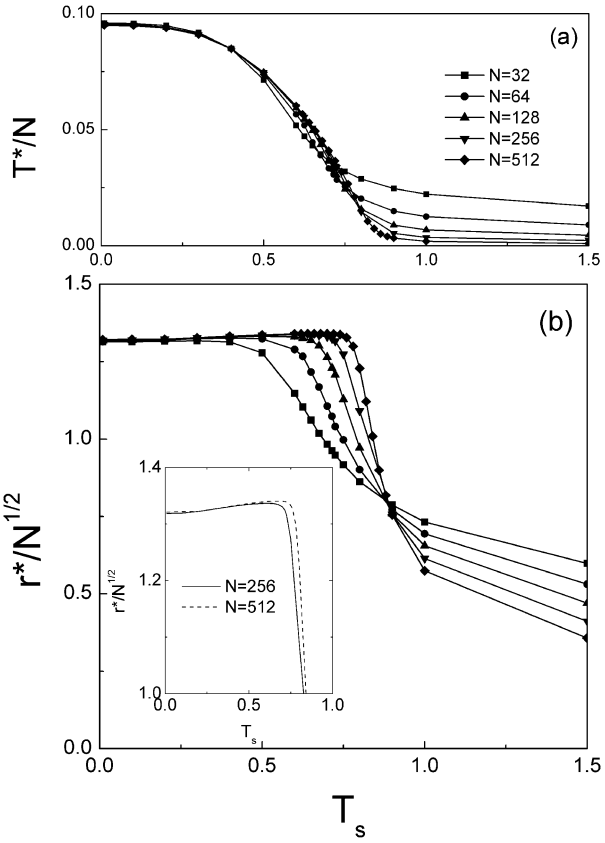

Fig. 35 **a** Transition temperature T^* and **b** domain size r^* ($= 2\pi/q^*$) as a function of the sequence design temperature for different chain lengths N

symmetric diblock copolymers, $T^*/N = 0.096$ or $\chi^* N = 10.4$). Therefore, it is interesting to explore a range of values for T_s that yields a compromise between these two regimes. As seen, there is a certain critical design temperature ($T_s^* \approx 0.78$) near which the value of r^* has a maximum for the designed sequences with $N \geq 256$.

Let us consider the thermodynamics of the transition in sequence space from high design temperatures, where many sequences contribute more or less equally, to the lowest design temperature, where one or a few sequences dominate. Figure 36 shows the distribution of domain sizes $W(r^*)$ for the ensemble of sequences generated near T_s^*. There is a clear bimodality seen in $W(r^*)$. Therefore, the transition in sequence space occurs as a first-order-like transition. The most important observation that can be made for this transitory regime is that the sequences providing the maximum of r^* do not correspond to simple symmetric diblocks but rather they look like gradient ones. Another intriguing result is that the Jensen–Shannon divergence

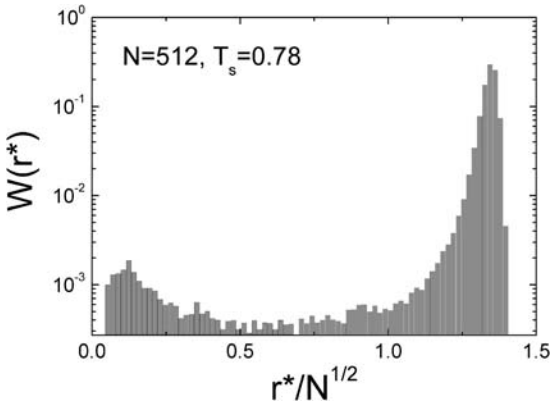

Fig. 36 Distribution of domain sizes $W(r^*)$ for the ensemble of sequences generated near T_s^*

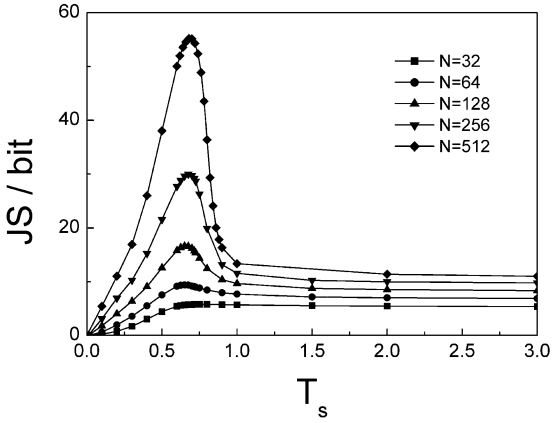

Fig. 37 The Jensen–Shannon divergence measure as a function of the sequence design temperature for different chain lengths N

measure shows a maximum in the vicinity of T_s^* (Fig. 37). This behavior is currently not completely understood, but might be a good starting point for further investigations.

3.3 Charged Hydrophobic Copolymers

Hydrophobic polyelectrolytes (HPEs) are polymers composed of covalently bonded sequences of polar (ionizable) groups which are soluble in water and hydrophobic groups which are not. This unique duality towards an aqueous environment leads to a rich spectrum of complex self-association phenom-

ena. Such polymers can also be considered as highly simplified models of biologically important macromolecules, e.g., proteins [96, 162].

3.3.1 Solution Properties

It is well known that many of the solution properties of proteins are due to a complex interplay between short-range hydrophobic attraction, long-range Coulomb effects, and the entropic degrees of freedom [163]. The balance of these factors determines the solubility and equilibrium conformation of proteins in water. To a large extent, these properties are also dependent on the unique primary structure of the polypeptide chain. This fact motivates intensive experimental and theoretical studies directed to the design of HPEs with specific chemical sequences resembling those of biomacromolecules. In addition, enormous effort has recently been directed toward the computer simulation of model HPEs with the purposes, explicit or implicit, of imitating the conformational behavior of biological macromolecules or of obtaining a basic understanding of the mechanism of molecular aggregation. In the present section we will focus on these studies.

Conformational transitions of single-chain HPEs are today part of an important theoretical challenge because of the very specific and somewhat unusual conformations they are supposed to adopt [164]. Much attention has been paid to the study of HPEs, mainly using theoretical concepts and coarse-grained simulation models [163]. It was found that in dilute salt-free HPE solutions, the subtle and antagonist balance of electrostatic and hydrophobic interactions can lead to a large variety of conformations compared to polyelectrolytes without hydrophobic groups. In particular, simulation results demonstrate that there is a range of electrolyte concentration and hydrophobicity for which HPEs exhibit exotic but stable conformations, namely the pearl necklace and the cigar-shaped conformations [163]. By gradually decreasing the monomer hydrophobicity of a strong polyelectrolyte, it undergoes a cascade of transitions from a nearly spherical globule to a cylinder-shaped conformation, a cylinder to a pearl necklace, and a pearl necklace to a strongly extended structure, successively [165–168]. These microstructures are controlled by a balance between surface tension and electrostatic free energies.

Micka et al. [169] were the first who simulated a multichain HPE system. They studied regular copolymers with alternating neutral and charged monomers (with a charge fraction of $f = 1/3$) in a poor solvent in the presence of monovalent counterions. The paper by Micka et al. [169] nicely demonstrated that the necklace microstructures exhibit a variety of conformational transitions as a function of polymer concentration. The end-to-end distance was found to be a nonmonotonic function of concentration and showed a strong minimum in the semidilute regime.

The solution properties of charged hydrophobic-hydrophilic protein-like copolymers have been studied in the presence of both mono- and multivalent counterions using molecular dynamics simulations [170–172]. In the model, one half of the polymer units was assumed to be negatively charged (P units) while the hydrophobic H units were electrically neutral but assumed to be strongly attractive. The variation of temperature allowed for the covering of a wide range of chain states from strongly swollen polyions at high temperature to strongly collapsed polyions at low temperature. The question of primary structure has also been explored, using microscopic polymer integral equation (RISM) theory [173]. The results of the simulations [170, 171] were found to be strongly dependent on the valence of the counterions and the temperature. The effect of these factors is discussed below.

3.3.2
Designed Copolymers in the Presence of Monovalent Counterions

The main conclusion drawn from the simulations [170] is that in the presence of monovalent counterions, the charged protein-like copolymers can be soluble, even in a very poor solvent for hydrophobic units. There are three temperature regimes, which are characterized by different spatial organization of polyions and their conformational behavior.

At sufficiently high temperature, i.e., in the weak coupling regime, the chains have extended coil-like or necklace-like conformations and are distributed more or less uniformly in the solution. Such a behavior is typical for single-chain HPEs [164].

At lower temperature, due to the intrachain hydrophobic association, there is a sharp decrease in chain size: the flexible-chain polyions rearrange in globules with neutral chain sections located in the globular core and charged sections forming the envelope of the core and buffering it from solvent. The collapsed globules, however, still have a net charge and repel each other. As a result, in this intermediate temperature regime, one observes a stable solution of nonaggregating polymer globules, which are well separated from each other and form an array of colloid-like particles with partly condensed counterions [170].

Due to the presence of microscopic and mesoscopic charged particles, two different length scales and large charge ratios are involved which make the mixed system strongly asymmetric. For better understanding this type of behavior, the following computer experiment was carried out [170]. First, two globules were "stuck" together, taking into account only hydrophobic attraction while the electrostatic interactions were turned off. The obtained nearly spherical globular agglomerate was relaxed for a long time. Then, the Coulomb interaction was turned on and the system evolution was monitored in the presence of counterions. The sequence of images, which illustrates the

evolution of the coalesced globules after switching on the Coulomb interaction, is shown in Fig. 38. The disintegration of the biglobular "droplet" into two separated globules is clearly seen, thereby indicating that the charged protein-like globules can indeed be stable in a poor solvent with respect to aggregation.

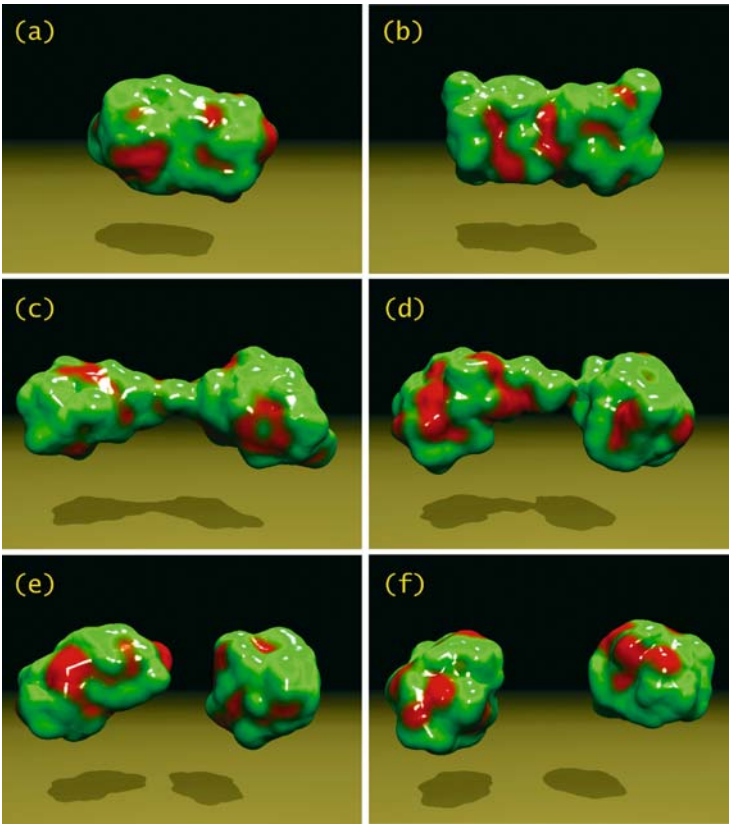

Fig. 38 **a–f** Sequence of images, which illustrates the evolution of the biglobular agglomerate after switching on the Coulomb interaction. For visual clarity, we present three-dimensional Connolly surfaces constructed for the polymer segments and do not depict the counterions. The repulsive Coulomb forces acting between non-compensated charges on the chains attempt to disrupt the polyion "droplet" into two "splinters", while the surface tension tries to keep it spherical. That situation is almost the same as that observed by Lord Rayleigh for highly charged liquid droplets. Depending on the droplet radius and surface tension (or temperature), the electrified droplet can become unstable, i.e., it starts to deform into an elongated ellipsoid. Such a deformation can finally lead to a fission of the droplet into two fragments of equal size and charge, if the repulsive force between the like elementary charges on the surface exceeds the forces from surface tension. Regions occupied by charged (hydrophilic) units are shown in *green* and hydrophobic regions are colored in *red*

A further decrease in temperature leads to the condensation of most of the counterions (the so-called ionomer regime) and to the aggregation of dense globules [170]. Nevertheless, they maintain their morphological integrity even in rather concentrated solutions where no large-scale aggregation is observed. Aggregated polyions form a specific supramolecular structure built up from stable individual globules connected together due to both short-range adhesive forces and effective mutual attractions mediated by counterions condensed on the strongly charged globule surfaces. In fact, the compact protein-like globules entering the aggregates behave as solid charged particles nonpenetrating each other. Interestingly, those finite size mesoglobular aggregates have an anisotropic chainlike morphology [170].

While simulations have demonstrated convincing evidence for effective attraction stemming from the counterion correlations, the physical mechanism underlying the onset of attraction is still not completely understood [163]. Also, it is not clear the role that details of the discrete charge distribution on the surface of the protein-like globules may play. In principle, this is relevant for any globular proteins with a nontrivial charge pattern on their surface. In view of this it has to be admitted that the traditional model of a homogeneously smeared charge [174] does not always seem to be correct.

3.3.3
Effect of Multivalent Counterions

When multivalent counterions ($z \geq 2$) are present in the solutions of charged protein-like macromolecules [171], the coil-to-globule transition and counterion condensation shift towards higher temperatures as compared to the $z = 1$ case [170]. Although condensed counterions of higher valence induce stronger interchain attraction and weaker repulsion, the aggregation of protein-like copolymers is not accompanied by large-scale aggregation. Instead, the formation of stable finite-size aggregates consisting of several entangled chains is observed. Charged chain segments tend to be on the surface of the aggregates, thereby screening their hydrophobic interior from the solvent. There is a penetration of the counterions into the aggregate core, although they show somewhat inhomogeneous distribution, tending to localize in outer aggregate regions. Even at low temperatures, there is no exact neutralization of negative and positive charges in the volume occupied by multichain aggregates. A certain fraction of counterions always remains in solution.

The average number of copolymer chains per aggregate, m, found in [171] is shown in Fig. 39 as a function of temperature for the systems containing counterions of different valence. It is seen that regardless of the valence of the counterion, the average aggregate size increases as the temperature is decreased. This is, of course, a quite expected result. One can conclude that for all the systems, the aggregation process becomes well pronounced in the

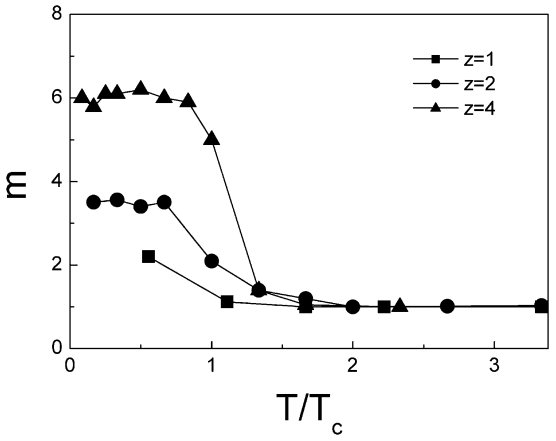

Fig. 39 Average number of 128-unit protein-like chains per aggregate as a function of reduced temperature T/T_c for counterions of different valence: ■ $z = 1$, ● $z = 2$, and ▲ $z = 4$. T_c is the temperature of counterion condensation. Adapted from [171]

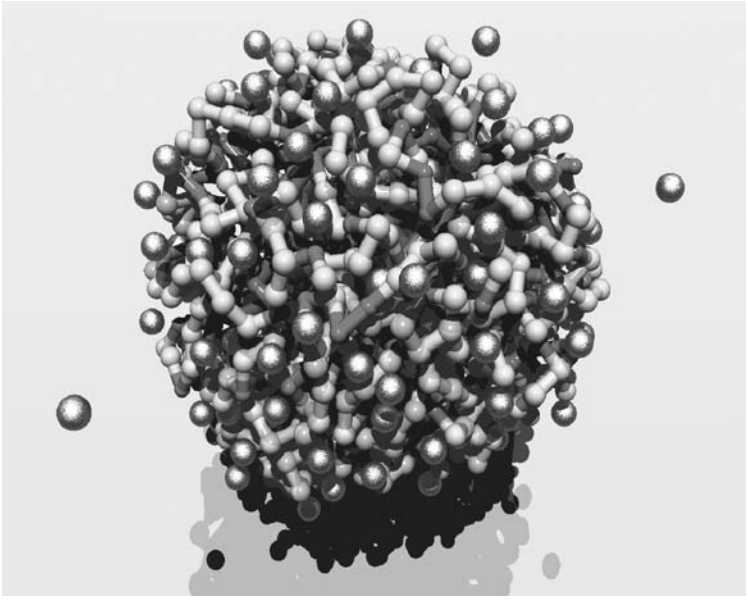

Fig. 40 Structure of a typical intermolecular aggregate formed in the presence of tetravalent counterions ($z = 4$). The aggregate is built up from seven 128-unit hydrophobic protein-like polyions, which are entangled, and 110 tetravalent counterions so that the net charge of the aggregate is – 8. It is seen that counterions condense preferably on the surface of the nearly spherical aggregate of entangled chains. The aggregate surface is covered with polar chain sections. A fraction of the counterions is floating in the immediate vicinity of the polyions. The charged chain groups are presented as *light gray spheres*, and the neutral chain groups are shown in *dark gray*. Counterions are shown as *larger spheres*

temperature region $T \leq T_c$, where T_c is the critical temperature of counterion condensation. In this case, the chains adopt compact conformations. Although the existence of separated single-chain globules is a prevalent structural motif for the $z = 1$ system [170], nevertheless, the occasional formation of intermolecular aggregates is observed at very low temperature. In the presence of di- and tetravalent ions, the average aggregate size is considerably larger. However, for all the systems studied in [171] the interchain aggregates had a finite size even at the lowest temperature; that is, no large-scale aggregation is observed. Figure 40 shows a typical snapshot of an intermolecular aggregate formed at $z = 4$.

3.3.4
Stabilization Mechanism

There are two driving forces causing the formation of compact conformations of charged copolymer chains: the short-range hydrophobic attraction and the counterion-mediated attraction between charged chain segments. In recent years, the origin of counterion-mediated attraction has been extensively discussed in the literature (see, e.g., [163]). About 10 years ago, Ray and Manning [175] suggested that two like-charged objects (e.g., rigid rods) can share condensed counterions in such a way to allow them to form what is analogous to a chemical bond. This "bridging type" model has been extensively used to explain the counterion-induced precipitation of polyelectrolytes [163, 175] and other related phenomena. When the temperature is reduced, the counterion condensation takes place that makes the counterion-mediated attraction stronger. Obviously, this effect is more pronounced for counterions with higher charge.

The stabilization of the finite-size aggregates (both single-chain and multichain) is due to the interplay between the attractive forces and the long-range Coulomb interactions together with the contribution from the translational entropy of mobile counterions. The point is that only a fraction of the charges on the polyions are neutralized by condensed counterions because the temperature is not zero, so that each copolymer chain still carries a net negative charge. Therefore, when the formation of the aggregates occurs, their net charge (i.e., the sum of the net charges of the individual chains in the aggregate) turns out to be nonzero for all aggregate sizes. The electrostatic repulsion due to the net-charge, whose range is set by the screening length under given conditions, is the *first factor* responsible for the stabilization of finite size aggregates. Obviously, the repulsion is longer in range than the attraction. Also, the simulations [170–172] show that for the HP copolymers having a protein-like primary structure, the significant fraction of charged chain segments is concentrated on the aggregate surface. Such a morphology assists the stabilization. On the other hand, it seems evident that this will not be the case for polyions with a purely random primary structure or for regu-

lar copolymers with an alternating distribution of charged and neutral groups along the chain.

In the counterion condensation regime, the remaining (noncondensed) counterions are partially immobilized near the charged monomer units and this results in translational entropy losses. This is the *second ingredient* of the stabilization mechanism.

For small clusters, the long-range Coulomb repulsion is more important, while for large interchain aggregates the stabilization should originate mainly from the counterion entropy decrease. Indeed, as the aggregates grow and occupy a larger fraction of the available volume in the system, a larger fraction of the free counterions will be found in the aggregates, thereby leading to an increase in the corresponding free energy connected with the translational entropy losses. Also, it should be noted that the counterion mobility inside an aggregate is reduced when the chains contract significantly and the aggregates become denser. All these arguments explain the existence of the finite-size aggregates built up from charged copolymers in a poor solvent.

The stabilization mechanism described above is practically the same as that suggested in [176]. In particular, the theory developed in this work predicts that in the dilute solution of associating polyelectrolytes, finite-size clusters having some optimum dimension should exist. Moreover, this stabilization mechanism is universal and should be valid for all attracting polyelectrolytes independent of the nature of the attraction between chain segments and the internal structure of the aggregates. To illustrate this assertion, the authors of [176] present the following qualitative arguments. When the polyions aggregate into a multichain cluster, the intermolecular potential energy of attraction per chain is a monotonically decreasing function of the aggregation number, m. It is clear that this energy should decrease up to some finite negative constant value for the macroscopic cluster in the $m \to \infty$ limit. On the other hand, the sum of the electrostatic repulsive energy and the counterion entropy decrease (per chain) is a monotonically increasing function of m. As a result, there should be a minimum of the total free energy at some finite value of m that corresponds to the formation of optimum finite-size aggregates for the given conditions. An analogous stabilization mechanism may come via the generalization of the approaches developed by Borue and Erukhimovich [177] and by Joanny and Leibler [178].

We can add a few words about the nonmonotonic behavior of the chain size observed for the low-temperature region [170, 171]. Let us compare the conformation of a polyion entering a single-chain and multichain aggregate. In the single-chain aggregate (i.e., in a dense polymer globule), every strongly collapsed polyion has a size of the order of $N^{1/3}$. In the large multichain aggregate, however, each flexible-chain macromolecule is entangled with other chains, which form a surrounding essentially similar to that characteristic of a polymer melt. In the $m \to \infty$ limit, each flexible chain should approximately behave as a Gaussian chain with the size of the order of $N^{1/2}$, i.e., its

size increases. That is why the chains included into larger aggregates formed in the presence of multivalent ions have larger sizes than those observed for small single-chain aggregates [171]. The chain expansion is also clearly seen in Fig. 40.

3.3.5 Experimental Results

Experimentally, there are some hints for the existence of finite-size HPE aggregates.

Carbajal-Tinoco et al. [179] have shown that even very hydrophobic sodium poly(styrene-*co*-styrene sulfonane) chains close to the solubility limit in water can be treated as completely isolated, colloid-like particles. Peng and Wu [180] have studied an aqueous solution of copolymers prepared by polymerization of *N*-vinylcaprolactam and sodium acrylate, P(VCL-*co*-NaA). Under poor solvent conditions in the presence of Na^+ ions, they observed a temperature-induced coil-to-globule transition of the chain conformation, leading to a single-chain core-shell nanostructure in which negatively charged COO^- groups are concentrated in the solution-exposed part of the globule. The average hydrodynamic radius and the average monomer density gradually decrease as the solvent became poorer, but neither the apparent weight-average molar mass and the average number of aggregation change, clearly indicating that there is no intermolecular aggregation. In the presence of Ca^{+2} ions, the intermolecular aggregates became larger, but their size was always finite, showing the existence of the stable mesoglobular phase. These results are in good qualitative agreement with the findings [170] for charged protein-like copolymers. The formation of the finite size clusters has also been observed for hydrophobically modified polyelectrolyte gels [181].

It is known that, depending on the primary amino acid sequence of proteins, there are two possible scenarios of protein aggregation in solution, when the attractive hydrophobic interaction dominates over the stabilizing electrostatic repulsion as the net charge of the protein is reduced. In the first case, globular proteins are able to form stable aggregates (dimers, trimers, etc.) maintaining in general their native conformations which they have in the low-concentration limit. The self-association and dimerization of lysozime, experimentally studied for a considerable time [182–184], is a typical example of this scenario. In the second case, proteins can also aggregate, but, in order to gain the maximum potential energy and to form the most stable structure, they should refold into a conformation dissimilar from the native (single-chain) state, the aggregation scenario suggested, e.g., for the cellular prion protein PrP [185, 186], capable of converting from the PrP^c form with α-helices into the PrP^{sc} misfolded structure having β-sheets. It may be assumed that the simple model of the charged protein-like copolymers in the presence of monovalent counterions under conditions corresponding to

the low-temperature regime reflects some general features of the aggregative processes observed for real proteins in the framework of the first scenario described above.

3.4 Hydrophobic-Amphiphilic Copolymers

The coil-to-globule transition in synthetic polymers, occurring under poor solvent conditions, 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, globular proteins differ from the globules of synthetic 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 monomeric units, we can formulate the following problem: is it possible to design such a sequence of an HP copolymer that provides globules protection from aggregation in solution?

Why is the understanding of the aggregation mechanism so important? First, it is needed for in vitro or computer design of new (soluble) protein-like copolymers. Second, it can help to gain insight into the stability of protein solutions. It is known that protein association leading to reduced biological activity plays a vital role in fundamental biological processes. In

particular, aggregation of the human proteins (e.g., lysozyme) that can form stable amyloid fibrils is associated with a range of fatal diseases including systemic amyloidosis, Alzheimer's disease, and transmissible spongiform encephalopathy [187]. Knowing which conformational rearrangements converge on the same final fold is important for understanding the determinants of protein structure, and may enable the development of rational approaches to the inhibition of protein condensation diseases. Simple isotropic models that treat the proteins as hard spherical colloids with short-range attractive interactions explain some features of the protein phase diagram. They, however, fail to describe the properties of protein solutions quantitatively and cannot address phenomena such as protein self-assembly [174, 188].

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 HP copolymer model. With this model, Abkevich et al. [160] 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 [189]. 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 [190]. Maximally compact conformations were found to be destabilized in solution as the intermolecular potential varies. Timoshenko and Kuznetsov [191] 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 [192, 193] 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 [194, 195]). Linear protein-like chains, capable of forming core-shell conformations, do not exhibit such a high tendency for aggregation as their random and random-block counterparts [40]. Nevertheless, they are not completely protected from aggregation [40], in contrast to many real protein globules. Therefore, it is instructive to look at other factors, which can be responsible for aggregation stability.

Real proteins are built up both from 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 can also determine some of the unique solution properties of globular proteins. We have already discussed the results of simulations

for charged protein-like hydrophobic-hydrophilic copolymers with a fixed charge distribution under poor solvent conditions [170, 171]. Indeed, for this model we have observed a solution of nonaggregating globular macroions. It should be kept in mind, however, that many of the nonaggregating proteins are uncharged.

Thus, we can conclude that until now, theoretical and experimental attempts to design nonaggregating heteropolymers have not given 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 electro-neutral heteropolymers, since in this case there are no real ways of creating the insurmountable (or very high) energy barriers, preventing large-scale aggregation. Nevertheless, we should note that for very long chains, certain irregular arrangements of H and P units along the chain may help to prevent precipitation of the copolymers thus leading to formation of stable finite aggregates and globules [196] (see also [197]).

Another method leading to nonaggregating copolymers may be connected with the molecular design of their monomeric units. We have discussed an extended variant of the HP model, the HA side-chain model [97], that explicitly takes into account the amphiphilic nature of hydrophilic segments.

It is believed that, using this model and the methods of sequence design suitable for computer simulations, one can construct copolymers capable of forming 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.

Although we are primarily interested in the intermolecular effects, we will first characterize the behavior of single amphiphilic copolymers in a selective solvent.

3.4.1 Single Amphiphilic Chains

Homopolymers consisting of amphiphilic monomer units (poly-A, Fig. 24a) were simulated, using a Langevin molecular dynamics method and the "side-chain" model [97]. The simulations of the hydrophobically driven conformational transitions under the variation of solvent conditions have shown that for this model, a variety of novel structures with high complexity are possible, depending on the interaction between hydrophobic (H) and hydrophilic (P) sites. Specifically, the thermodynamically stable anisometric structures have been observed, including disk-like structures, stretched necklace-like conformations, and cylindrical-shaped conformations. Also, it was demonstrated that the chain size R_g as a function of the quality of the solvent can behave in an irregular manner, showing an *increase* when the solvent becomes poorer

for hydrophobic sites. This unusual behavior is connected with the formation of strongly elongated core-shell conformations having a locally cylindrical symmetry and is consistent with existing experimental data [198]. Moreover, for the range of the chain lengths N simulated ($N \leq 1024$), the formation of such conformations can lead to the $R_g \propto N^{0.9}$ scaling under a poor solvent condition.

In order to collect information on the qualitative features of the chain conformations, we can directly look at many snapshots of the amphiphilic chain when the solvent quality is progressive worsened. Figure 41 shows a series of typical snapshots obtained for the chain with a 256-unit backbone at the strong H – P segregation.

When the repulsive interactions between monomers dominate, the chain has the usual coil-like conformation (Figs. 41a and b). As the solvent becomes poorer, we observe chain folding and this leads to the formation of specific necklace-like conformations where single "pearls" of hydrophobic groups surrounded by hydrophilic groups are connected by stretched chain sections (Figs. 41c and d). In this regime, the mobility of each monomer is quite high and monomer position fluctuations are still large. Pearls are locally in equilibrium, linked to one another by fluctuating chain sections. With worsening solvent quality the size of pearls increases and, as a result, their number decreases. Finally, for very strong attraction between H sites, the pearls coalesce and form an object that looks like a sausage, which then transforms to a cylindrical-shaped object with the cross-section increasing

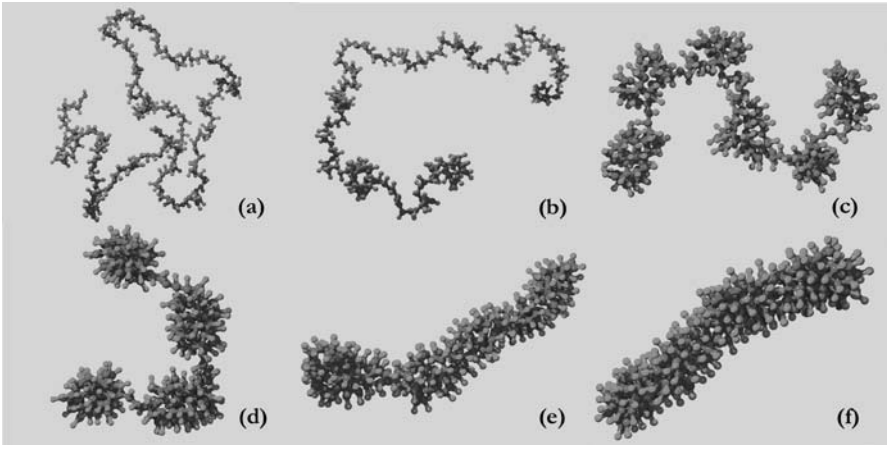

Fig. 41 Snapshot pictures illustrating typical conformations of the amphiphilic chain (poly-A) of length $N = 256$ for the strong H-P segregation, at different H – H attraction increasing from **a** to **f**. Hydrophobic beads are shown as *dark gray spheres* and hydrophilic beads are presented as *light gray spheres*. The sizes of all the spheres are schematic rather than space filling

very slowly as the H – H attraction grows (Figs. 41e and f). Such strongly nonspherical aggregates, whose state is liquid-like, are thermodynamically stable. Upon visual inspection of these aggregates, one can find that the hydrophobic chain finds itself in an irregularly folded (crumpled) state. Thus, the progressive worsening of solvent quality results in the following succession of the conformational transitions at sufficiently strong H – P segregation: swollen coil → stretched necklace-like conformation → sausage-like object → cylindrical-shaped conformation. We would like to emphasize that such transitions should be considered rather as a smooth shape evolution of the model polymer than sharp transitions. In other words, there is no well-defined temperature transition between different regimes but rather a smooth transition from the necklace-like regime to the sausage regime and further to the cylindrical regime. Therefore, it is possible to observe, e.g., a coexistence of pearls and sausage for a given solvent condition.

It is important to note that the observed 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: each of the hydrophobic (hydrophilic) groups has a tendency to have hydrophobic (hydrophilic) nearest neighbors and to avoid having hydrophilic (hydrophobic) nearest neighbors in a poor solvent. 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. It is clear that such a situation is very similar to that characteristic of usual low-molecular-weight amphiphiles, which form micelles in a dilute solution [199, 200]. Indeed, the "side-chain" model [97] bears resemblance to a system of N small chemically connected HP surfactants. In a polar solvent, surfactants can form micelles with a dense hydrophobic core surrounded by a hydrophilic shell. Also, such a behavior is rather common for polysoaps which, due to their amphiphilic character, are able to build up intramolecular self-assembled structures in polar as well as in apolar media [201–205]. From this viewpoint, we should treat the conformational transitions found for the extended HP model (Fig. 24a) rather as an intrachain micellization than as a true coil-to-globule transition. In particular, the necklace-like conformation with hydrophobic pearls surrounded by hydrophilic groups (Figs. 41c and d) has to be considered as a string of micelles but it has nothing in common with the pearl-necklace model structure proposed by Rubinstein and co-workers [164, 206] for flexible polyelectrolytes in poor solvents where, due to the Rayleigh charge instability, highly stretched segments alternate with collapsed (micro)globules along the chain. Structurally, the intramolecular micelles observed for the necklace-like state are similar to micelles formed by free low-molecular-weight surfactants; however, unlike ordinary micelles, intramolecular micelles need no critical concentration of polysurfactants for their formation because in this case there is no loss of translational entropy. To understand

many of the results of the computer simulations [97], simple theoretical arguments [207–210] that predict the formation of nonspherical micelles can be used.

The conformational transitions observed in the simulations [97] resemble in some aspects the so-called *zipping* transitions [211], the process in which two strongly attracting strands composing the polymer come in contact in such a way as to form a bound double structure, which remains swollen and does not assume compact configurations. The cylindrical-shaped conformations in which the hydrophobic backbone is in a locally collapsed state (Figs. 41e and f) look a lot like three-dimensional zipped structures.

Single hydrophobic-amphiphilic (HA) copolymers with the same HA composition but with different distribution of H and A units along the main hydrophobic chain were also simulated [212]. In particular, regular copolymers comprising H and A units in alternating sequence, regular multiblock copolymers composed of H and A blocks of equal lengths, and the quasirandom protein-like copolymers with a quenched primary structure were studied. These copolymers are schematically depicted in Fig. 24b,c, and d.

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. It was found that single protein-like polyamphiphiles (Fig. 24d) 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 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 ($N \leq 255$), 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.

The globules of relatively short regular multiblock copolymers with a fixed block length of $L = 3$ (Fig. 24c) 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 the alternating HA sequence (Fig. 24b). 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. For sufficiently long chains, the formation of such conformations leads to the $R_g \propto N^{\nu}$ scaling with $0.86 \leq \nu \leq 0.89$. This scaling exponent is practically the same as that observed for poly-A chains. The fact that the scaling exponent is slightly smaller than the rod-like prediction of $\nu = 1$ is likely due to the fact that N is very close to, though not yet in, the scaling regime. Nevertheless, it is believed that in the $N \to \infty$ limit, because of strong

thermal fluctuations, the locally folded chain as a whole would look like an infinitely coiled "garden hose" (or a worm-like superchain) having a finite thickness and a finite persistent length, and the limiting scaling exponent would be close to that expected for the good solvent regime, due to repulsive interactions between the outer hydrophilic groups. In other words, for very large N we expect a crossover from the $R_g \propto N$ regime to the $R_g \propto N^{0.59}$ regime.

There are experimental evidences of some facts predicted in the simulations [97, 212].

Kikuchi and Nose [198, 213] have reported on systematic experimental studies of poly(methylmethacrylate)-*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 [213] have observed the formation of thermodynamically stable unimolecular rod-like 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 [214] have demonstrated that poly(styrene)-*graft*-poly(4-vinyl-*N*-ethylpyridium bromide) forms unimolecular micelles in water/methanol mixtures. These experimental data can be treated as an indirect confirmation of the simulation result [97] 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 a solution of amphiphilic copolymers has also been observed by Wu and Qiu [215]. 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 a collapsed PNIPAM chain backbone as the hydrophobic core and the grafted short PEO chains as the hydrophilic shell. In general, these colloid-like nanostructures are similar to those observed in the simulations [212] for protein-like amphiphilic copolymers.

In a series of papers [216, 217], Nakata and Nakagawa have studied the coil-globule transition by static light scattering measurements on poly(methyl methacrylate) in a selective solvent. They have found that the chain expansion factor, $\alpha^2 = R_g^2/R_{g\Theta}^2$, plotted against the reduced temperature, $\tau = 1 - \Theta/T$, first decreases with decreasing τ, as it should be, but then begins to increase (see, e.g., Fig. 2 presented in [217]) In the authors opinion, "the increase of

α^2 with decreasing temperature conflicts with theoretical predictions and an intuitive notion of the expansion factor." However, taking into account the amphiphilic nature (although weakly-pronounced) of poly(methyl methacrylate) and the simulation data reported for the extended HP model [97], this "anomalous" behavior can be understood. Indeed, in solvent selective to side groups, the incompatibility of chemically different groups is effectively increased when the attraction between groups composing the chain backbone becomes stronger. This is accompanied by pushing away the soluble side groups from the insoluble micellar core and by stretching the macromolecule as a whole.

Williams and co-workers [179, 218] have studied the structural changes and chain conformations of a series of hydrophobic sodium poly(styrene-*co*-styrene sulfonate)'s of various charge fractions in a poor solvent using static light scattering and small-angle X-ray scattering techniques. By varying the charged monomer fraction, f, it was possible to change the degree of hydrophobicity of this copolymer and the corresponding hydrophobic/hydrophilic interactions. From the scattering measurements, the so-called apparent radii of gyration, R_g^{app}, were determined for different values of f and concentrations. From the analysis of these results (see Tables 1 and 2 of [179]) it is seen that the values of R_g^{app} show an irregular behavior as a function of f at all the polymer concentrations studied. At large f, when the repulsive electrostatic forces dominate, the copolymer chains have an expanded conformation. With decreasing f, the intrachain short-range hydrophobic attractions begin to dominate and, as a result, the chain size decreases. However, at $f \lesssim 1/2$, an unexpected increase in R_g^{app} is distinctly observed; although, at first sight, R_g^{app} should further decrease, taking into account a monotonous growth in hydrophobic attraction. It is clear, that such an "unexpected" behavior is consistent with the simulations [97, 212] and can be explained on the basis of the discussion presented above.

3.4.2
Coil-Globule Transition Versus Aggregation

Here, we describe and compare the results of simulations for two multichain systems corresponding to alternating and protein-like HA copolymers [212]. The multichain systems consisting of 127-unit copolymers were simulated for the range of the effective interaction parameter $\widetilde{\chi}$ (which is similar to the Flory–Huggins parameter) under solvent conditions when single chains can form strongly collapsed conformations.

In Fig. 42, we show the ratio R_{gP}^2/R_{gH}^2 (R_{gH}^2 and R_{gP}^2 are the partial mean-square radii of gyration calculated separately for hydrophobic and hydrophilic beads) as a function of the interaction parameter $\widetilde{\chi}$. We see that this ratio is an increasing function of $\widetilde{\chi}$. Qualitatively the same picture is observed for isolated chains. Such behavior is due to the fact that, as the attraction

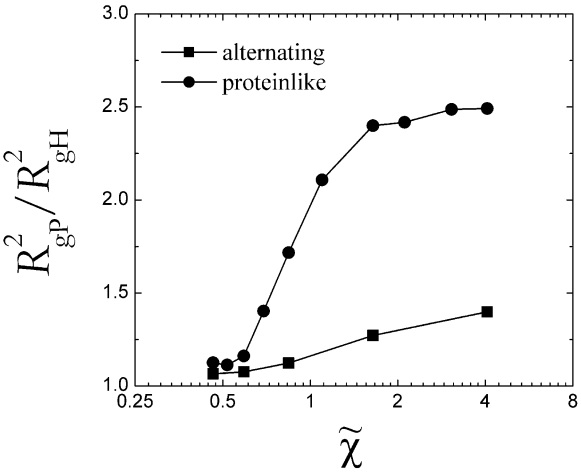

Fig. 42 Ratio R^2_{gP}/R^2_{gH} as a function of the interaction parameter $\widetilde{\chi}$ in a semi-logarithmic scale for the 127-unit ■ alternating and • protein-like chains in the corresponding multichain systems. The parameter $\widetilde{\chi}$ is similar to the Flory–Huggins interaction parameter χ and characterizes solvent quality in an integral manner. Sufficiently large values of $\widetilde{\chi}$ ($\widetilde{\chi} \gtrsim 1$) correspond to a poor solvent. Solvent quality becomes poorer with decreasing temperature or with increasing $\widetilde{\chi}$. Adapted from [212]

between H segments increases, the value of R^2_{gP} decreases more slowly than R^2_{gH}, thus leading to demixing of H and P segments and facilitating their intramolecular microphase separation. This trend is more pronounced for the protein-like copolymers than for the alternating copolymers, suggesting an idea that the former should be more protected against intermolecular aggregation.

A direct way to study the process of aggregation is to monitor the change in the corresponding free energy, ΔG. To this end, following the standard quasi-chemical approach, one 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 pair correlation function; 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. 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 Fig. 43 as a function of the interaction parameter $\widetilde{\chi}$. As seen, lowering the temperature, or equivalently,

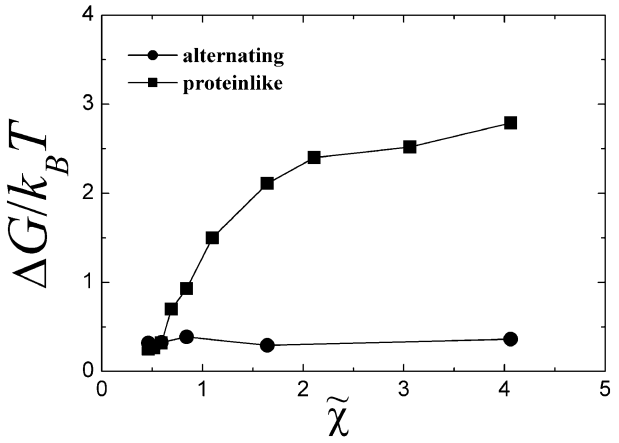

Fig. 43 Aggregation free energy ΔG as a function of the interaction parameter $\tilde{\chi}$ for the multichain systems of (■) alternating and (●) protein-like copolymers. Adapted from [212]

increasing the interaction parameter $\tilde{\chi}$, shifts the equilibrium $A_1 \rightleftarrows A_M$ ($M \geq 2$) in the system of protein-like copolymers toward the nonaggregated state, and this is reflected in Fig. 43 as an increase in ΔG.

At first sight, such behavior is somewhat counterintuitive. Actually, from a general consideration it would be possible to expect the 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 a 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 a strengthening of this hydrophilic "protective barrier". On 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 $\tilde{\chi}$. This behavior also can be understood on the basis of the data concerning the conformational structure of this copolymer. Since at the same HA composition the cylindrical globules of alternating copolymer 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 protein-like copolymers rather high densities of the hydrophilic shell are reached.

Moreover, the thorough analysis of globular conformations shows that this layer is almost absent near the faces of the cylinder [212]. This facilitates the formation of multiglobular aggregates in the solution of regular copolymers.

These facts explain the more expressed tendency of regular copolymers toward aggregation.

For visual analysis of the simulated configurations, one can employ the technique based on the construction of isosurfaces. In this way, the global system morphology can be studied. The snapshots of the low-temperature configurations ($\widetilde{\chi} \approx 4$) 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 (Fig. 44). In this respect, the behavior observed for the model amphiphilic copolymers is similar to that found for charged hydrophobic/hydrophilic protein-like chains [170–172]. The simulation [212] 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, one observes a stable solution of nonaggregating polymer globules which are well-separated from each other and form an array of colloid-like 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.

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. In the HA model, "surfactants", i.e., amphiphilic A groups, are incorporated into the polymer chain, thus ensuring the stabilizing effect. From the tempera-

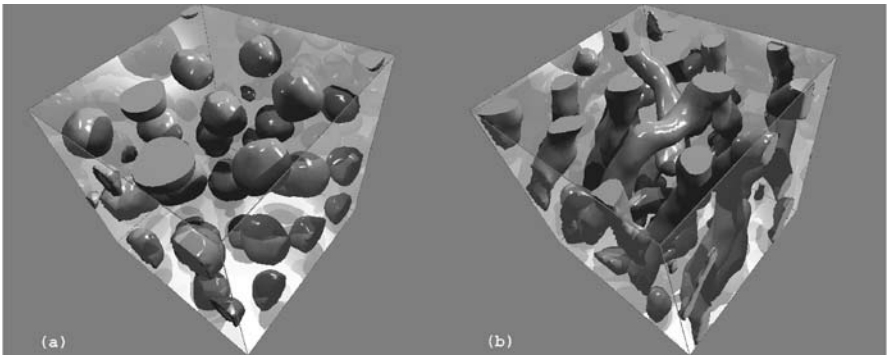

Fig. 44 Snapshot pictures representing the isosurfaces generated under poor solvent conditions for the multichain systems composed of the 127-unit amphiphilic copolymers with **a** protein-like and **b** alternating distribution of H and A groups along the chain

ture dependencies of aggregation free energy (Fig. 43) one can conclude that below the collapse transition temperature there is a 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.

The situation with the alternating copolymers is not so clear. Although in this case the large-scale aggregation is also not observed [212], such copolymers were found to be capable of forming sufficiently large intermolecular aggregates, as seen in Fig. 44.

3.5
Adsorption Selectivity

The focus here will be on the consideration of designed copolymers exhibiting selective interactions with the surfaces and interfaces. In particular, we will consider some properties of adsorption-tuned copolymers and partly cross-linked polymer envelopes that function as a molecular dispenser.

3.5.1
Adsorption-Tuned Copolymers

The adsorption of homopolymers at an impenetrable surface is a well-studied problem [219, 220]. Much less is known about copolymer adsorption (in which only one of two comonomers interacts with the surface), although the problem has been studied by several groups [221–225]. Regular copolymers with a periodic sequence of comonomers, adsorbing at a planar surface, have been studied by Moghaddam et al. [226]. In the case where the copolymer is random, the most interesting case is *quenched randomness* where the sequence of comonomers is fixed during the computation of thermodynamic quantities, which are averaged over the quenched comonomer sequences. Grosberg et al. [227] have considered copolymers with periodic quenched randomness. The case of nonperiodic quenched randomness has also been studied using a variety of techniques [228–230]. Moghaddam and Whittington [231] have used multiple Markov chain Monte Carlo methods to investigate the adsorption of a random copolymer at a homogeneous surface. Adsorption of an ideal correlated random copolymer at a liquid-liquid interface has been investigated theoretically in [232].

Zheligovskaya et al. [55] have simulated the adsorption of quasirandom adsorption-tuned copolymers (ATC). The critical adsorption energy as well as some characteristics of the adsorbed single chains (statistics of trains, loops, and tails) were studied. All these properties were compared with those

of random copolymers with the same content of adsorbed segments and random-block copolymers with the same composition and the same average numbers of adsorbed and nonadsorbed blocks.

It was found that the difference in the primary structure of the chains leads to the difference in the critical adsorption energy ε^* and the characteristics of adsorbed chains. In particular, the ATC chains have the smallest (by the absolute value) adsorption energy ε^*. The random copolymers are characterized by the largest ε^*, as compared to other copolymers. This fact is simply explained by the difference in block lengths: the random copolymers have the shortest blocks. This is consistent with the analytical results [233] for regular AB copolymers, according to which the absolute value of the critical adsorption energy decreases with the increasing block length at the same fraction of adsorbed and nonadsorbed segments. At the same time, the difference in the critical adsorption energy for the random-block and ATC chains (which are characterized by the same average block lengths) can be explained only by the details of the ATC primary structure. It turns out to be that the ATC chains have significantly longer end nonadsorbed blocks, as compared to their random-block counterparts. As a result, the adsorbed segments are placed more compactly in each ATC chain. This specific feature of the ATC primary structure promotes adsorption of ATC chains. The studied characteristics of the adsorbed chains are also different for the three copolymers.

Thus, the difference in the adsorption behavior of ATC, random, and random-block chains can be rationally explained by taking into account the specific features of the primary structure of these chains.

The obtained results support the general idea of a conformation-dependent sequence design of copolymers proposed [18–20]. That is, the generated ATC sequence memorizes some features of the specific parent conformation of the adsorbed homopolymer. In particular, the position of adsorbed segments turned out to be tuned in the best way for subsequent adsorption. It is not surprising, therefore, that this memorized hidden information became apparent as soon as we considered the adsorption of ATC chains. Among the three types of AB copolymers which were studied [55], ATC chains adsorb better at a given adsorption energy. In other words, the AB chain "learns to be adsorbed" in the parent conformation, and this "experience" is used in the subsequent "life" of this copolymer.

3.5.2 Molecular Dispenser

To characterize the complexes formed between molecular dispenser described in Sect. 2.2.4 and colloidal particles, the probability $P(\sigma, T)$ of finding a complex made from the copolymer envelope and the particle of a given size, σ, was calculated as a function of temperature T [57].

After the preparation of the copolymer envelope (Fig. 10), the parent particle of size σ_p was eliminated, and another particle of a given diameter σ was introduced into the system. A new particle was placed far from the copolymer envelope; that is, initially it was not interacting with the envelope. During the stochastic motion in the bulk, the particle and envelope came into collision with each other, thus forming a copolymer-particle complex. Under the thermal agitation the complex was splitting and reassembling again. The function $P(\sigma, T)$, which is related to the energy of interaction between the copolymer envelope and particle, was calculated as an average over $\sim 10^2$ independent realization.

It was found that the selectivity of the complex formation strongly depends on the number of crosslinks in the envelope. Typical results for a moderately crosslinked envelope are shown in Fig. 45a. It is seen that the selectivity of the complex formation with the particle of a certain size is indeed reached, that is, the idea of a molecular dispenser works.

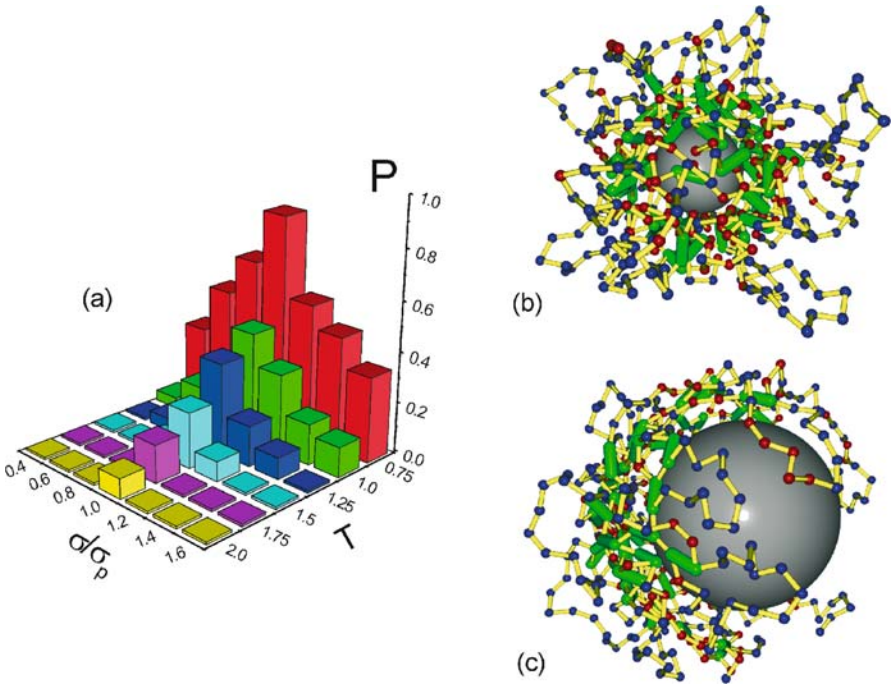

Fig. 45 **a** Probability of finding a complex made from a 512-unit copolymer envelope and a particle of a given size, σ, at the temperature T for the case when the copolymer envelope has 48 crosslinks. Snapshots of the complexes made from a 512-unit copolymer envelope for **b** $\sigma/\sigma_p = 0.8$ and **c** $\sigma/\sigma_p = 1.8$, where σ_p is the size of the parent particle. Adsorbed and non-adsorbed chain segments are colored in *red* and *blue*, respectively; crosslinks are shown as *green sticks*. Adapted from [57]

Generally, the structure of the polymer-particle complex can be found from the minimization of free energy that includes the polymer-particle interaction energy, entropies of nonadsorbed monomer units and units localized at the surface of the particle, and typically, for the system under consideration, the elastic deformation of crosslinked macromolecule. Such theoretical analysis, following the lines of [234, 235], can explain the specific behavior of $P(T, \sigma)$ observed for the envelopes with different numbers of crosslinks n_J [57]. According to [235], when the number of crosslinks is small enough, $n_J \ll N^{1/2}$, all junctions contribute mainly to the formation of simple loops along the chain and the polymer molecule as a whole conserves linear structure. On the other hand, when $n_J > N^{1/2}$, the macromolecule becomes really crosslinked, forming a kind of loose network. In the simulation [57], the cross-links introduced between adsorbed monomeric units stabilize a hollow-spherical structure of the copolymer envelope with mesh-like architecture and the cage structure of the central cavity. Thus, the selective property of the moderately crosslinked copolymer envelope is based on the existence of a certain mesh size, which restricts the penetration of big particles into the copolymer envelope.

The reason for the selective adsorption of a colloidal particle of parent size is explained by the typical snapshots in Figs. 45b and c. We see that the particle of parent or smaller size ($\sigma \leq \sigma_p$) is fully absorbed by the central cavity (Fig. 45b), because the corresponding fitting was ensured by the sequence design procedure (Fig. 10b). On the other hand, a particle of larger size ($\sigma > \sigma_p$) turns out to be too big for a central cavity (Fig. 45c), and thus the complex formed does not saturate all the possibilities for the attraction of red units to the surface of the particle. As to small particles, they easily penetrate inside the molecular dispenser, but the complex formed is not stable (especially at high temperature) because of the small surface of such particles. All these factors explain the peak in $P(T, \sigma)$ observed at $\sigma \approx \sigma_p$ for moderately crosslinked copolymer envelopes (Fig. 45a).

4 Conclusion

In this review, we reported on several new synthetic strategies that allow the synthesis of copolymers with a broad variation of their sequence distributions. The fundamental principle of these strategies is conformational-dependent sequence design (CDSD), which takes into account a strong coupling between the conformation and primary structure of copolymers during their synthesis. Using computer simulation techniques, we have attempted to show that rather simple methods, such as polymer-analogous reactions and normal radical copolymerization, can lead to nontrivial chemical sequences, long-range correlations, and gradient structures, if they take place under unusual physical conditions.

The results presented in this review demonstrate that the CDSD polymer-analogous transformation is a versatile approach that allows various functional copolymers such as bioinspired protein-like macromolecules (which give soluble globules with a segregated core-shell microstructure), molecular dispensers (which are able to selectively absorb nanoparticles of a given size), and adsorption-tuned copolymers to be obtained. Our discussion has focused on copolymer sequences exhibiting large-scale compositional heterogeneities and long-range statistical correlations between monomeric units. These features are intrinsically related to the CDSD scheme and they cannot be explained by the basic stochastic processes such as random sequence or Markov chain; the first has no correlations and the second only has short-range correlations. Problems associated with the evolution of copolymer sequences have been considered from the viewpoint of the emergence of information complexity in the sequences in the course of evolution.

Presently the main technique for the synthesis of copolymers are free-radical polymerization methods [11–17]. For a limited range of comonomers, anionic and cationic polymerizations are also used [236, 237].

We have reviewed results on the computer modeling of radical copolymerization under heterogeneous conditions, including the following synthetic methodologies: solution copolymerization with simultaneous globule formation, emulsion polymerization with polymerizing hydrophobic and amphiphilic monomers, copolymerization near a chemically homogeneous surface that selectively adsorbs one type of polymerizing monomer, and template copolymerization near patterned surfaces. Because the implementation of conventional radical polymerization is much easier than that of anionic or CRP processes, it is worthwhile to try to use the emerging features of the copolymer sequences (long-range correlations and gradient structures) to design new types of copolymers with sophisticated functional properties. Many of these advances are likely to lead soon to novel applications.

Also, we have discussed advances that have recently been achieved in the computer simulation and theoretical understanding of designed copolymers in solution and in bulk. The focus was on amphiphilic protein-like copolymers and on hydrophobic polyelectrolytes. Here, we have tried to demonstrate how the copolymer sequence dictates the structure and properties of polymer systems. In many cases, the presence of long-range correlations in designed copolymers can bring about dramatic changes in their physical properties with respect to the corresponding copolymers whose sequence has only minor correlation between adjacent monomeric units.

Acknowledgements Many of our colleagues have contributed to these studies. In particular, we are grateful to A.V. Berezkin, A.Yu. Grosberg, V.A. Ivanov, S.I. Kuchanov, P. Reineker, and A.N. Semenov whose ideas and/or encouragement were critical for some of the questions discussed in this work. The financial support from the Alexander von Humboldt Foundation, Program for Investment in the Future (ZIP), SFB 569, NWO, and RFBR (project # 04-03-32185) is highly appreciated.

References

1. Hamley IW (1998) The physics of block copolymers. Oxford University Press, Oxford
2. Blumenfeld LA, Tikhonov AN (1994) Biophysical thermodynamics of intracellular processes: molecular machines of the living cell. Springer, Berlin Heidelberg New York
3. Li H, Helling R, Tang C, Wingreen N (1996) Science 273:666
4. Xia Y, Levitt M (2004) Curr Opin Struct Biol 14:202
5. Shakhnovich EI (1996) Folding & Design 1:R50
6. Shakhnovich EI (1997) Curr Opin Struct Biol 7:29
7. Shakhnovich EI (1998) Folding & Design 3:45
8. Frauenfelder H, Wolynes PG, Austin RH (1999) Rev Mod Phys 71:S419
9. Pande VS, Grosberg AY, Tanaka T (2000) Rev Mod Phys 72:259
10. Konig JL (1980) Chemical microstructure of polymer chains. Wiley, New York
11. Greszta D, Madrare D, Matyjaszewski (1994) Macromolecules 27:638
12. Moad G, Solomon DH (1995) The chemistry of free radical polymerization. Elsevier, Oxford
13. Malmstrom EE, Hawker CJ (1998) Macromol Chem Phys 199:923
14. Clay P, Christie DI, Gilbert RG (1998) Controlled radical polymerization. In: Matyjaszewski K (ed) ACS Symposium Series 685. Am Chem Soc, Washington, DC, Chap. 7
15. Matyjaszewski K, Xia J (2001) Chem Rev 101:2921
16. Davis KA, Matyjaszewski K (2002) Adv Polym Sci 159:14
17. Davis KA, Matyjaszewski K (2002) Statistical, gradient, block and graft copolymers by controlled/living radical polymerizations. Springer, Berlin Heidelberg New York
18. Khokhlov AR, Khalatur PG (1998) Physica A 249:253
19. Khalatur PG, Ivanov VI, Shusharina NP, Khokhlov AR (1998) Russ Chem Bull 47:855
20. Khokhlov AR, Khalatur PG (1999) Phys Rev Lett 82:3456
21. Khokhlov AR, Grosberg AY, Khalatur PG, Ivanov VA, Govorun EN, Chertovich AV, Lazutin AA (2001) In: Broglia RA, Shakhnovich EI (eds) Proceedings of the International School of Physics *Enricon Fermi*, Course CXLV: Protein folding, evolution and design. IOS Press, Amsterdam, p 313–332
22. Khokhlov AR, Khalatur PG, Ivanov VA, Chertovich AV, Lazutin AA (2002) In: Mac Kernan D (ed) Challenges in molecular simulations (SIMU Newsletter). CECAM, Lyon, vol 4, p 79–100
23. Khokhlov AR, Khalatur PG (2004) Curr Opin Solid State Mater Sci 8:3
24. Khokhlov AR, Berezkin AV, Khalatur PG (2004) J Polym Sci A Polym Chem 42:5339
25. Khalatur PG, Berezkin AV, Khokhlov AR (2004) Recent Res Develop Chem Phys 5:339
26. Shakhnovich EI, Gutin AM (1993) Proc Natl Acad Sci USA 90:7195
27. Pande VS, Grosberg AY, Tanaka T (1994) J Phys France 4:1771
28. Pande VS, Grosberg AY, Tanaka T (1994) Proc Natl Acad Sci USA 91:12972
29. Irbäck A, Peterson C, Potthast F, Sandelin E (1998) Phys Rev E 58:R5249
30. Gupta P, Hall CK, Voegler AC (1998) Protein Sci 7:2642
31. Lau KF, Dill KA (1989) Macromolecules 22:3986
32. Hamley IW (ed) (2004) Developments in block copolymer science and technology. Wiley, New York
33. Schultz GE, Schirmer RH (1979) Principles of protein structure. Springer, Berlin Heidelberg New York
34. Carmesin I, Kremer K (1988) Macromolecules 21:2819

35. Govorun EN, Ivanov VA, Khokhlov AR, Khalatur PG, Borovinsky AL, Grosberg AY (2001) Phys Rev E 64:040903
36. Shlesinger MF, Zaslavskii GM, Frisch U (1996) Lévy flights and related topics in physics. Springer, Berlin Heidelberg New York
37. Kuchanov SI, Khokhlov AR (2003) J Chem Phys 118:4684
38. Virtanen J, Baron C, Tenhu H (2000) Macromolecules 33:336
39. Virtanen J, Tenhu H (2000) Macromolecules 33:5970
40. Virtanen J (2002) Self-assembling of thermally responsive block and graft copolymers in aqueous solutions (PhD Thesis). Department of Chemistry, University of Helsinki, Helsinki, p 42
41. Li W, Kaneko K (1992) Europhys Lett 17:655
42. Fink TMA, Ball RC (2001) Phys Rev Lett 87:198103
43. Peng C-K, Buldyrev SV, Goldberger AL, Havlin S, Sciortino F, Simon M, Stanley HE (1992) Nature (London) 356:168
44. Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Phys Rev E 49:1685
45. Pande VS, Grosberg AY, Tanaka T (1994) J Chem Phys 101:8246
46. Irbäck A, Peterson C, Potthast F (1996) Proc Natl Acad Sci USA 93:9533
47. Gusev LV, Vasilevskaya VV, Makeev VY, Khalatur PG, Khokhlov AR (2003) Macromol Theory Simul 12:604
48. Yu ZG, Anh VV, Lau KS (2003) Phys Rev E 68:021913
49. Irbäck A, Peterson C, Potthast F (1996) Proc Natl Acad Sci USA 93:9533
50. Yu ZG, Anh VV, Lau KS (2001) Phys Rev E 64:031903
51. Doukhan P, Oppenheim G, Taqqu MS (eds) (2003) Theory and applications of long-range dependence. Birkhäuser, Boston
52. Khalatur PG, Berezkin AV, Khokhlov AR (2005) J Chem Phys (submitted)
53. Kuchanov SI (2000) Adv Polym Sci 152:157
54. Wesson L, Eisenberg D (1992) Protein Sci 1:227
55. Zheligovskaya EA, Khalatur PG, Khokhlov AR (1999) Phys Rev E 59:3071
56. Starovoitova NY, Khalatur PG, Khokhlov AR (2004) In: forces, growth and form in soft condensed matter: at the interface between physics and biology. In: Skjeltorp AT, Belushkin AV (eds) NATO Science Series II: Mathematics, Physics and Chemistry. Kluwer, Dordrecht, vol 160, p 253
57. Velichko YS, Khalatur PG, Khokhlov AR (2003) Macromolecules 36:5047
58. Thurmond KB, Kowalewski T, Wooley KL (1997) J Am Chem Soc 119:6656
59. Huang H, Remsen EE, Kowalewski T, Wooley KL (1999) J Am Chem Soc 121:3805
60. Sanji T, Nakatsuka Y, Ohnishi S, Sakurai H (2000) Macromolecules 33:8524
61. Stewart S, Liu G (1999) Chem Mater 11:1048
62. Discher BM, Won Y-Y, Ege DS, Lee JCM, Bates FS, Discher DE, Hammer DA (1999) Science 284:1143
63. Nardin C, Hirt T, Leukel J, Meier W (2000) Langmuir 16:1035
64. Trefonas P, West R, Miller RD (1985) J Am Chem Soc 107:2737
65. Smith JM (1972) On evolution. Edinburgh Univ Press, Edinburgh
66. Volkenstein MV (1983) General biophysics. Academic Press, New York
67. Leninger AL, Nelson DL, Cox MM (1993) Principles of biochemistry. 2nd edn, Worth Publishers, New York
68. Grosberg AY, Khokhlov AR (1997) Giant molecules: Here and there and everywhere. Academic Press, New York
69. Gatlin LL (1972) Information theory and the living system. Columbia Univ Press, New York
70. Khalatur PG, Novikov VV, Khokhlov AR (2003) Phys Rev E 67:051901

71. Chertovich AV, Govorun EN, Ivanov VA, Khalatur PG, Khokhlov AR (2004) Eur Phys J E 13:15
72. Grosberg AY (1984) Biofizika 29:569
73. Wullf G, Sarhan A (1972) Angew Chem Int Ed 11:341
74. Polowinski S (1997) Template polymerization. ChemTec Publishing, Toronto Scarborough
75. de Gennes P-G (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca New York
76. Berezkin AV, Khalatur PG, Khokhlov AR (2003) J Chem Phys 118:8049
77. Berezkin AV, Khalatur PG, Khokhlov AR, Reineker P (2004) New J Phys 6:44
78. Berezkin AV, Khalatur PG, Khokhlov AR (2005) Polymer Sci A 47:66
79. Flory PJ (1953) Principles of polymer chemistry. Cornell University Press, Ithaca New York
80. Lozinskii VI, Simenel IA, Kurskaya EA, Kulakova VK, Grinberg VY, Dubovik AS, Galaev IY, Mattiasson B, Khokhlov AR (2000) Dokl Chem 375:273
81. Lozinsky VI, Simenel IA, Kulakova VK, Kurskaya EA, Babushkina TA, Klimova TP, Burova TV, Dubovik AS, Grinberg VY, Galaev IY, Mattiasson B, Khokhlov AR (2003) Macromolecules 36:7308
82. Wahlund P-O, Galaev IY, Kazakov SA, Lozinsky VI, Mattiasson B (2002) Macromol Biosci 2:33
83. Siu M-H, Zhang G, Wu C (2002) Macromolecules 35:2723
84. Siu M-H, He C, Wu C (2003) Macromolecules 36:6588
85. Dotson NA, Galvan R, Laurence RL, M Tirrell (1996) Polymerization process modeling. Wiley, New York
86. Berezkin AV, Khalatur PG, Khokhlov AR (2005) Polymer Sci (submitted)
87. Starovoitova NY, Khalatur PG, Khokhlov AR (2003) Dokl Chem 392:242
88. Starovoitova NY, Berezkin AV, Kriksin YA, Gallyamova OV, Khalatur PG, Khokhlov AR (2005) Macromolecules 38:2419
89. Chan HS, Dill KA (1991) Annu Rev Biophys Chem 20:447
90. Peppas NA, Huang Y (2002) Pharmaceutical Res 19:578
91. Muthukumar M (1995) J Chem Phys 103:4723
92. Golumbfskie AJ, Pande VS, Chakraborty AK (1999) Proc Natl Acad Sci USA 96:11707
93. Kriksin YA, Khalatur PG, Khokhlov AR (2005) J Chem Phys 122:114703; Kriksin YA, Khalatur PG, Khokhlov AR (2005) Math Model 17:3
94. Berezkin AV, Solov'ev MA, Khalatur PG, Khokhlov AR (2004) J Chem Phys 121:6011
95. Berezkin AV, Solov'ev MA, Khalatur PG, Khokhlov AR (2005) Polymer Sci A 47:622
96. Alexandridis P, Lindman B (eds) (2000) Amphiphilic block copolymers: self-assembly and applications. Elsevier, Amsterdam
97. Vasilevskaya VV, Khalatur PG, Khokhlov AR (2003) Macromolecules 36:10103
98. Kriksin YA, Khalatur PG, Khokhlov AR (2003) Macromol Symp 201:29
99. Banavar JR, Maritan A (2003) Rev Modern Phys 75:23
100. Ivanov VA, Chertovich AV, Lazutin AA, Shusharina NP, Khalatur PG, Khokhlov AR (1999) Macromol Symp 146:259
101. van den Oever JMP, Leermakers FAM, Fleer GJ, Ivanov VA, Shusharina NP, Khokhlov AR, Khalatur PG (2002) Phys Rev E 65:041708
102. Ptitsyn OB (1992) In: Creighton TE, Freeman WH (eds) Protein folding. W.H Freeman, New York, p 253–300
103. Kriksin YA, Khalatur PG, Khokhlov AR (2002) Macromol Theory Simul 11:213
104. Nishio I, Sun S-T, Swislow G, Tanaka T (1979) Nature 281:208
105. Swislow G, Sun S-T, Nishio I, Tanaka T (1980) Phys Rev Lett 44:796
106. Chu B, Ying Q (1996) Macromolecules 29:1824

107. Wu C, Zhou S (1996) Phys Rev Lett 77:3053
108. Chu B, Ying Q, Grosberg AY (1995) Macromolecules 28:180
109. Nakata M, Nakagawa T (1997) Phys Rev E 56:3338
110. Wu C, Wang X (1998) Phys Rev Lett 80:4092
111. Zhang G, Wu C (2001) Phys Rev Lett 86:822
112. von Rague-Schleyer P, Allinger NL, Clark TC, Gasteiger J, Kollman PA, Schaefer HF (1998) Encyclopedia of computational chemistry. Wiley, New York
113. Rost B, Sander C (1995) Protein structure prediction by neural networks. In: Arbib M (ed) The Handbook of brain theory and naural networks. The MIT press, Cambridge, MA, p 772–775
114. de Gennes P-G (1975) J Phys Lett 36:L55
115. Buguin A, Brochard-Wyart F, de Gennes P-G (1996) C R Acad Sci Paris 322:741
116. Ostrovsky B, Bar-Yam Y (1994) Europhys Lett 25:409
117. Abrams CF, Lee N-K, Obukhov S (2002) Europhys Lett 59:391
118. ten Wolde PR, Chandler D (2002) Proc Natl Acad Sci USA 99:6539
119. Lee N-K, Abrams CF (2004) J Chem Phys 121:7484
120. Polson JM, Moore NE (2005) J Chem Phys 112:024905
121. Yue K, Dill KA (1992) Proc Natl Acad Sci USA 89:4163
122. Cooke IR, Williams DRM (2003) Macromolecules 36:2149
123. Bates FS, Fredrickson GH (1999) Phys Today 52:32
124. Fredrickson GH, Bates FS (1996) Annu Rev Mater Sci 26:501
125. Bates FS, Fredrickson GH (1996) Annu Rev Phys Chem 41:525
126. Khalatur PG (2000) Polymer Sci C 42:229
127. Ikkala O, ten Brinke G (2004) Chem Commun 2131
128. Hajduk DA, Harper PE, Gruner SM, Honeker CC, Kim G, Thomas EL, Fetters LJ (1994) Macromolecules 27:4063
129. Förster S, Khandpur AK, Zhao J, Bates FS, Hamley IW, Ryan AJ, Bras W (1994) Macromolecules 27:6922
130. Goldacker T, Abetz V, Stadler R, Erukhimovich I, Leibler L (1999) Nature 398:137
131. Nap R, Erukhimovich I, ten Brinke G (2004) Macromolecules 37:4296
132. Leibler L (1980) Macromolecules 13:1602
133. Helfand E, Wasserman ZR (1982) In: Goodman I (ed) Developments in block copolymers. Applied Science, London, p 99
134. Matsen MW, Schick M (1994) Phys Rev Lett 72:2660
135. Matsen MW, Schick M (1996) Curr Opin Colloid Interface Sci 1:329
136. Matsen MW, Bates FS (1996) Macromolecules 29:1091
137. Matsen MW (2001) J Phys Condens Matter 14:R21
138. Drolet F, Fredrickson GH (1999) Phys Rev Lett 83:4317
139. Fredrickson GH (2002) J Chem Phys 117:6810
140. Schweizer KS, Curro JG (1994) Adv Polym Sci 116:319
141. Schweizer KS, Curro JG (1997) Adv Chem Phys 98:1
142. David EF, Schweizer KS (1994) J Chem Phys 100:7767
143. David EF, Schweizer KS (1994) J Chem Phys 100:7784
144. Chandler D, Andersen HC (1972) J Chem Phys 57:1930
145. Chandler D, McCoy JD, Singer SJ (1986) J Chem Phys 85:5977
146. Gutin AM, Sfatos CD, Shakhnovich EI (1994) J Phys A 27:7957
147. Angerman H, ten Brinke G, Erukhimovich IY (1996) Macromolecules 29:3255
148. Angerman H, ten Brinke G, Erukhimovich IY (1996) Macromol Symp 112:199
149. Potemkin II, Panyukov SV (1998) Phys Rev E 57:6902
150. Semenov AN (1997) J Phys II (France) 7:1489
151. Semenov AN, Likhtman AE (1998) Macromolecules 31:9058

152. Semenov AN (1999) Eur Phys J B 10:497
153. Zherenkova LV, Talitskikh SK, Khalatur PG, Khokhlov AR (2002) Dokl Phys Chem 382:23
154. Zherenkova LV, Khalatur PG, Khokhlov AR (2003) Dokl Phys Chem 393:293
155. Fredrickson GH, Ganesan V, Drolet F (2002) Macromolecules 35:16
156. Swift BW, Olvera de la Cruz M (1996) Europhys Lett 35:487
157. Houdayer J, Müller M (2002) Europhys Lett 58:660
158. Dobrynin AV, Erukhimovich IY (1995) J Phys I (France) 5:365
159. Gutin AM, Abkevich VI, Shakhnovich EI (1995) Proc Natl Acad Sci USA 92:1282
160. Abkevich VI, Gutin AM, Shakhnovich EI (1996) Proc Natl Acad Sci USA 93:839
161. Gusev LV, Khalatur PG, Khokhlov AR (in preparation)
162. Branden C, Tooze J (1991) Introduction to protein structure. Garland, New York
163. Holm C, Joanny JP, Kremer K, Netz RR, Reineker P, Seidel C, Vilgis TA, Winkler RG (2004) Adv Polym Sci 166:67
164. Rubinstein M, Dobrynin AV (1999) Curr Opin Colloid Interface Sci 4:83
165. Chodanowski P, Stoll S (1999) J Chem Phys 111:6069
166. Limbach HJ, Holm C, Kremer K (2002) Europhys Lett 60:566
167. Lee N, Thirumalai D (2001) Macromolecules 34:3446
168. Chang R, Yethiraj A (2003) J Chem Phys 118:6634
169. Micka U, Holm C, Kremer K (1999) Langmuir 15:4033
170. Khalatur PG, Khokhlov AR, Mologin DA, Reineker P (2003) J Chem Phys 119:1232
171. Mologin DA, Khalatur PG, Khokhlov AR, Reineker P (2004) New J Phys 6:133
172. Khokhlov AR, Khalatur PG (2005) Curr Opin Colloid Interface Sci 10:22
173. Zherenkova LV, Khalatur PG, Khokhlov AR (2003) J Chem Phys 119:6959
174. Braun FN (2002) J Chem Phys 116:6826
175. Ray J, Manning GS (1994) Langmuir 10:2450
176. Potemkin II, Vasilevskaya VV, Khokhlov AR (1999) J Chem Phys 111:2809
177. Borue VY, Erukhimovich IY (1988) Macromolecules 21:3240
178. Joanny J-P, Leibler L (1990) J Phys (France) 51:545
179. Carbajal-Tinoco MD, Ober R, Dolbnya I, Bras W, Williams CE (2002) J Phys Chem B 106:12165
180. Peng S, Wu C (2001) J Phys Chem B 105:2331
181. Philippova OE, Andreeva AS, Khokhlov AR, Islamov AK, Kukin AI, Gordeliy VI (2003) Langmuir 19:7240
182. Sophianopoulos AJ, Holde KEV (1964) J Biol Chem 239:2516
183. Wang F, Hayter J, Wilson LJ (1996) Acta Crystallogr D 52:901
184. Booth DR et al. (1997) Nature 385:787
185. Prusiner SB (1991) Science 252:1515
186. Cohen FE, Pan K-M, Huang Z, Baldwin M, Fletterick R, Prusiner SB (1994) Science 264:530
187. Booth DR, Sundetll M, Bellotti V, Robinso CV, Hutchinson WL, Fraser PE, Hawkins PN, Dobson CM, Radfordt SE, Blaket CCF, Pepys MB (1997) Nature 385:787
188. Lomakin A, Asherie N, Benedek GB (1999) Proc Natl Acad Sci USA 96:9465
189. Broglia RA, Tiana G, Pasquali S, Roman HE, Vigezzi E (1998) Proc Natl Acad Sci USA 95:12930
190. Giugliarelli G, Micheletti C, Banavar JR, Maritan A (2000) J Chem Phys 113:5072
191. Timoshenko EG, Kuznetsov YA (2000) J Chem Phys 112:8163
192. Bratko D, Blanch HW (2001) J Chem Phys 114:561
193. Bratko D, Blanch HW (2003) J Chem Phys 118:5185
194. Gupta P, Hall CK, Voegler AC (1998) Protein Sci 7:2642
195. Toma L, Toma S (2000) Biomacromolecules 1:232

196. Semenov AN (2004) Macromolecules 31:226
197. Govorun EN, Khokhlov AR, Semenov AN (2003) Eur Phys J E 12:255
198. Kikuchi A, Nose T (1996) Macromolecules 29:6770
199. Mittal KL (ed) (1977) Micellization, solubilization, and microemulsions. Plenum Press, New York, London, vol 1 and 2
200. Smit B (1993) Computer simulations of surfactants. In: Allen MP, Tildesley DJ (eds) Computer simulation in chemical physics. Kluwer Academic Publishers, Dordrecht, p 461–472
201. Laschewsky A (1995) Adv Polym Sci 124:1 and references cited herein
202. Borisov OV, Halperin A (1995) Langmuir 11:2911
203. Borisov OV, Halperin A (1996) Europhys Lett 34:657
204. Borisov OV, Halperin A (1996) Macromolecules 29:2612
205. Zhou SQ, Chu B (2000) Adv Mater 12:545
206. Dobrynin AV, Rubinstein M, Obukhov SP (1996) Macromolecules 29:2974
207. Eriksson JC, Ljunggren S (1990) Langmuir 6:895
208. Khalatur PG, Khokhlov AR, Nyrkova IA, Semenov AN (1996) Macromol Theory Simul 5:713
209. Khalatur PG, Khokhlov AR, Nyrkova IA, Semenov AN (1996) Macromol Theory Simul 5:749
210. Borisov OV, Zhulina EB (2005) Macromolecules 38:2506
211. Baiesi M, Carlon E, Orlandini E, Stella AL (2001) Phys Rev E 63:041801
212. Vasilevskaya VV, Klochkov AA, Lazutin AA, Khalatur PG, Khokhlov AR (2004) Macromolecules 37:5444
213. Kikuchi A, Nose T (1996) Polymer 37:5889
214. Selb J, Gallot Y (1981) Makromol Chem 182:1491,1513,1775
215. Wu C, Qiu X (1998) Phys Rev Lett 80:620
216. Nakata M, Nakagawa T (1997) Phys Rev E 56:33838
217. Nakata M, Nakagawa T (1999) J Chem Phys 110:2703
218. Carbajal-Tinoco MD, Williams CE (2000) Europhys Lett 52:284
219. Eisenriegler E (1993) Polymers near surface. World Science, Singapore
220. Netz RR, Andelman D (2003) Phys Rep 380:1
221. Cosgrove T, Finch NA, Webster JRP (1990) Macromolecules 23:1334
222. Joanny J-F (1994) J Phys II (France) 4:1281
223. Sommer JU, Daoud M (1995) Europhys Lett 32:407
224. Whittington SG (1998) J Phys A Math Gen 31:3769
225. Chakraborty AK (2001) Phys Pep 342:259
226. Moghaddam MS, Vrbova T, Whittington SG (2000) J Phys A Math Gen 33:4573
227. Grosberg A, Izrailev S, Nechaev S (1994) Phys Rev E 50:1912
228. Garel T, Huse DA, Leibler S, Orland H (1989) Europhys Lett 8:9
229. Gutman L, Chakraborty AK (1994) J Chem Phys 101:10074
230. Orlandini E, Tesi MC, Whittington SG (1999) J Phys A Math Gen 32:469
231. Moghaddam MS, Whittington SG (2002) J Phys A Math Gen 35:33
232. Denesyuka NA, Erukhimovich IY (2000) J Chem Phys 113:3894
233. Zhulina EB, Skvortsov AM, Birshtein TM (1981) Vysokomol Soed A 23:304
234. Erukhimovich IY (1978) Vysokomol Soed B 20:10
235. Panyukov SV, Potemkin II (1997) J Phys I (France) 7:273
236. Hsieh HL, Quirk RP (1996) Anionic polymerization: principles and practical applications. Marcel Dekker, New York
237. Hillmyer M (1999) Curr Opin Solid State Matter Sci 4:559