Phase diagram of a Gaussian random copolymer

E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson

Theory and Computation Group, Centre for Colloid Science and Biomaterials, Department of Chemistry,
University College Dublin, Dublin 4, Ireland
(Received 15 November 1996)

We study the stationary limit of the self-consistent kinetic equations derived earlier by us for a model of a Gaussian random copolymer. The phase diagram of the model contains five different states separated by the collapse, glass, and folding "transitions." We perform a finite-size chain analysis and find that the kinetic accessibility of the folded state is strongly impeded for sufficiently long chains. For the high-density globule we propose a simplified treatment in terms of the three main order parameters. This approximation yields a qualitatively correct phase diagram and allows us to demonstrate the thermodynamic stability of the folded state. [S1063-651X(97)08005-7]

PACS number(s): 36.20.−r, 87.15.By

I. INTRODUCTION

Studying the conformational transitions in polymers possessing quenched random amphiphilicity is important for many applications in both synthetic and biological macromolecular systems [1–3]. There is a view in the scientific community that elucidation of the conformational states of simple models of proteins, along with their attendant kinetic laws, would be a crucial step in understanding protein folding and misfolding [2]. One expression of the protein folding problem is to determine how a one-dimensional primary sequence of amino acid residues relates to the three-dimensional structure of the folded protein, and to deduce the kinetics of folding process that brings a statistical ensemble of extended coils into an essentially unique native state. Much is written about this fundamental problem and we refer the reader to many excellent monographs, reviews, and original works on protein folding (see Refs. [4–9], and references therein). We mention this by way of motivation. However, we need not enlarge the discussion here in our rather short paper devoted to the coarse-grained statistical mechanical description of a Gaussian random copolymer.

In our previous work in Ref. [10] we have proposed a nonequilibrium Gaussian self-consistent method for studying kinetics of a model based on the Edwards-type effective free energy functional with quenched disorder in the monomer two-body interactions. The method has the advantages that it is valid for kinetics as well as for equilibrium, and that it is suitable for description of both the fractal coil and the condensed globular phases. In this sense it is a unique approach, though it does have its own weakness to which we shall come in Sec. III.

Previous numerical analysis of the self-consistent equations resulted in what we believe is the correct picture of kinetic folding pathway pertinent to the general features discovered experimentally in a number of proteins and discussed in Refs. [11,12]. Significantly, the method predicted the existence of a nonfully compacted kinetic intermediate, possessing frozen and partially misfolded structure with a significant number of hydrophobic units exposed on the exterior of the globule.

By taking the stationary limit of the kinetic equations we obtain a set of the equilibrium equations for determination of the free energy minima. It is interesting to note that this approach is an alternative to a more standard replica spin-glass method [13,14]. In fact, in Sec. II C of Ref. [10] we have shown that these equilibrium equations are the minima conditions of the variational free energy functional in a version of the Gibbs-Bogoliubov approach. We choose the trial Hamiltonian dependent on the disorder variables and explicitly average the variational free energy functional over the quenched disorder. Because we average the free energy rather than the partition function, there is no need to apply the replica trick. In doing so, however, we have to use an approximation that is justified only for sufficiently small dispersions of disorder Δ. Thus it would be of interest to compare some of our results with those of more standard replica calculations. We have done so in Ref. [15], which addresses similar issues by a replica variational approach, and found a fair agreement between the two approaches in a range of Δ. We have not been able to reach higher values of the dispersion of disorder in the replica formalism due to technical difficulties in numerical solution of many-dimensional minimax problem there. The present approach allows us to eliminate such a limitation and to study the whole phase diagram of the system.

At this point it is important to emphasize that the kinetics after a quench from a homopolymerlike initial state automatically preserves perturbativeness of the solution. Determination of the equilibrium transition curves is a more delicate problem. The cause of our concern is quite clear. Application of additional closure relations for the higher-order correlation functions may, in principle, violate the variational bound for the free energy in the Gibbs-Bogoliubov scheme. The error incurred thereby is of order Δ^4 in our case. Thus we may expect the current method to be inaccurate in predicting the transitions curves if the free energy undergoes a change only in the order Δ^4. Nevertheless, the kinetic information, such as position of the spinodal curves, should, in principle, remain reliable.

Here we address the equilibrium issues of the problem in the framework of our method. First of all, we study the complete phase diagram and consider the behavior of the relevant observables in different phases. This information is comple-

mentary to the results of Ref. [10]. Furthermore, we analyze the spinodal and transition curves of the folding transition for different chain lengths. This elucidates the finite-size dependence of the folding transition. Later on we propose a simplified ansatz in terms of the system size, the phase separation, and glass order parameters that can be obtained from the complete set of equations in the limit of the dense globule. We show that this gives a qualitatively correct description of the three dense globular states. As an important development, this approximation confirms the thermodynamic stability of the folded state for large chain lengths. Finally, we consider the possibility of improving the equilibrium free energy by including higher-order corrections renormalizing the second virial coefficient. We find that there is a unique value of the fourth-order correction coefficient α_4 that allows one to obtain the freezing transition curve.

II. THE SELF-CONSISTENT EQUATIONS

In this section we shall briefly define the model and write down the self-consistent equations derived in Ref. [10]. We also introduce necessary notations and define the important observables. For further explanations and details we refer the reader to Ref. [10].

It is convenient to use the Fourier transforms x_q of the monomer coordinates X_m in the chain index (see the Appendix). Note that we use letters of different case in order to distinguish between the two sets of coordinates. Description of kinetics of the model is based upon the Langevin equation,

$$\zeta \frac{d}{dt} \mathbf{x}_q(t) = -\frac{\partial H}{\partial \mathbf{x}_{-q}} + \boldsymbol{\eta}_q(t), \qquad (1)$$

$$\langle \eta_q^\alpha(t) \eta_{q'}^{\alpha'}(t') \rangle = 2k_B T \zeta \delta_{q+q',0} \delta^{\alpha\alpha'} \delta(t-t'), \qquad (2)$$

where $\zeta = N\zeta_b$, ζ_b is the bare friction constant, and N is the degree of polymerization. For the purpose of this paper we may disregard the hydrodynamic effect for it only affects time-dependent characteristics.

The model accounts for the connectivity of the chain, excluded volume effects, and the random amphiphilicity of the monomers. We choose the Edwards-type effective free energy functional, $H = \bar{H} + H_{\text{dis}}$, containing the homopolymeric $\bar{H}$ and the disordered H_{dis} terms, respectively,

$$\bar{H} = \frac{\kappa}{2} \sum_n (\mathbf{X}_{n+1} - \mathbf{X}_n)^2 + \sum_{L>2} \bar{u}_L \sum_{\{m\}} \prod_{i=1}^{L-1} \delta(\mathbf{X}_{m_i} - \mathbf{X}_{m_{i+1}}), \qquad (3)$$

$$H_{\text{dis}} = \frac{1}{2} \sum_{m_1 m_2} (\Lambda_{m_1} + \Lambda_{m_2}) \delta(\mathbf{X}_{m_1} - \mathbf{X}_{m_2}). \qquad (4)$$

Here κ is the connectivity constant, $\bar{u}_L$ are the virial coefficients of the excluded volume interactions, and summation over $\{m\}$ includes all values of indices $m_1, \ldots, m_L$ with $m_i \neq m_{i+1}$.

It is worthwhile to comment here on the origin of the term (4) of the effective free energy functional that is linear in the disorder variables [16]. One usually proceeds from the effective free energy functional,

$$H_{\text{ms}} = H_{\text{sol}}[\mathbf{R}_a] + H_{\text{mon}}[\mathbf{X}_n] - \sum_{n,a} I_n \delta(\mathbf{X}_n - \mathbf{R}_a),$$

which includes the terms describing the solvent degrees of freedom $\mathbf{R}_a$, the monomer degrees of freedom $\mathbf{X}_n$, e.g., in the form of Eq. (3), and a contact monomer-solvent interaction, characterized by the nth monomer hydrophobic strengths I_n, respectively. A simple way of deriving Eq. (4), proposed by Garel and Orland [1], would be then to explicitly use the solution incompressibility condition,

$$\rho_{\text{mon}}(\mathbf{y}) + \rho_{\text{sol}}(\mathbf{y}) = \sum_n \delta(\mathbf{y} - \mathbf{X}_n) + \sum_a \delta(\mathbf{y} - \mathbf{R}_a)$$

$$= \rho_0 = \text{const}, \qquad (5)$$

in order to integrate out the solvent degrees of freedom. This yields the partition function $Z_{\text{ms}} = Z_{\text{sol}}Z$, where the effect of the solvent influence on the monomer degrees of freedom appears in Z only via the following term in the effective free energy functional:

$$H = \sum_{n \neq n'} [u_2 + \tfrac{1}{2}(I_n + I_{n'})] \delta(\mathbf{X}_n - \mathbf{X}_{n'}) + \cdots. \qquad (6)$$

Now, by introducing $\bar{u}_2 = u_2 + I$ and $\Lambda_n = I_n - I$, where I is the mean value of I_n, we obtain Eq. (4).

The random variables Λ_m, and consequently their Fourier transforms λ_q, are assumed to possess a Gaussian distribution with the second momentum,

$$\overline{\lambda_q \lambda_{q'}} = \bar{\Delta}^2 \delta_{q+q',0}, \quad \bar{\Delta}^2 \equiv \Delta^2/N, \qquad (7)$$

where Δ has the meaning of the dispersion of disorder. Here and throughout we use the angular brackets $\langle A \rangle$ to denote the statistical averages over the noise and initial ensemble of monomer positions $\{\mathbf{x}(t=0)\}$ and the overbar $\bar{A}$ to denote averages over the quenched distribution of disorder $\{\Lambda\}$.

In our previous work in Ref. [10] we have derived in some approximation closed kinetic equations for the two types of correlation functions. Thus let us introduce the mean-squared amplitudes of the normal modes,

$$\mathcal{F}_q(t) \equiv \overline{F_q(t)}, \quad F_q(t) = \tfrac{1}{3} \langle |\mathbf{x}_q|^2(t) \rangle, \qquad (8)$$

and the disorder correlation functions,

$$\varphi_{qp}(t) \equiv \overline{\phi_{qp}(t)}, \quad \phi_{qp}(t) = \tfrac{1}{3} \lambda_{q-p} \langle \mathbf{x}_{-q}(t) \mathbf{x}_p(t) \rangle. \qquad (9)$$

These satisfy the following self-consistent equations:

$$\frac{\zeta}{2} \frac{d}{dt} \mathcal{F}_q(t) = -\frac{2}{3} \left(\mathcal{F}_q \frac{\partial \mathcal{A}}{\partial \mathcal{F}_q} + \sum_p \varphi_{qp} \frac{\partial \mathcal{A}}{\partial \varphi_{qp}} \right), \qquad (10)$$

$$\zeta \frac{d}{dt} \varphi_{qp}(t) = -\frac{2}{3} \left[\varphi_{qp} \left(\frac{\partial \mathcal{A}}{\partial \mathcal{F}_q} + \frac{\partial \mathcal{A}}{\partial \mathcal{F}_p} \right) + \bar{\Delta}^2 (\mathcal{F}_q + \mathcal{F}_p) \frac{\partial \mathcal{A}}{\partial \varphi_{qp}} \right], \qquad (11)$$

where $\mathcal{A}$ is the variational free energy functional [17]. This form of the kinetic equations has a transparent meaning. Indeed, the kinetics could be understood as a motion representing the flow of the whole statistical ensemble in the phase
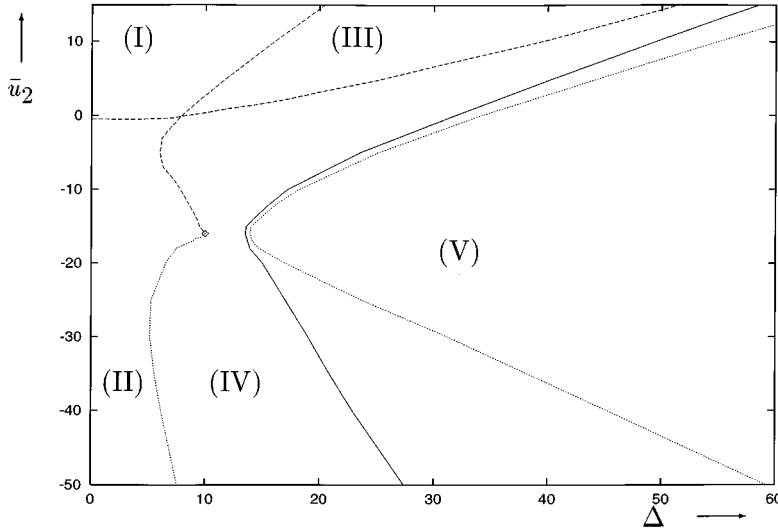
FIG. 1. The phase diagram of the model in terms of the second virial coefficient, $\bar{u}_2$ (in units $k_B T \mathcal{L}^3$), and the dispersion of disorder, Δ (in units $k_B T \mathcal{L}^3$). Solid lines represent first-order-like transitions, dashed lines represent continuous transitions, and dotted lines represent "spinodal" curves. The roman numerals correspond consequently to Flory coil, liquidlike globule, random coil, "glassy" phase, and folded globule. Continuous transition curves are determined by the points of the fastest change of respective order parameters ($\overline{R_g^2}$ for the collapse transition and $\overline{R_g^2 R_g^{2(c)}}$ for the glass transition). Here and below $N=30$, and $\bar{u}_3=10$ (in units $k_B T \mathcal{L}^6$).

space of the averaged dynamic variables. The motion is governed by the gradients and is directed towards the global free energy minimum.

The variational free energy $\mathcal{A}=\mathcal{E}-T\mathcal{S}$ contains the "entropic" part

$$S=k_B \frac{3}{2} \sum_q \log \mathcal{F}_q - k_B \frac{3\bar{\Delta}^{-2}}{4} \sum_{qp} \frac{\varphi_{qp}^2}{\mathcal{F}_q \mathcal{F}_p} + O(\bar{\Delta}^4), \quad (12)$$

and the mean energy $\mathcal{E}=\overline{\langle H \rangle}$,

$$\frac{\mathcal{E}}{N} = \frac{3\kappa}{2} \mathcal{D}_{01} + \hat{u}_2 \sum_k \frac{1}{\mathcal{D}_k^{3/2}} + \hat{u}_3 \sum_{k_1 k_2} \frac{1}{Y_0(k_1,k_2)^{3/2}}$$
$$- \frac{3}{2} \hat{1} \sum_k \frac{\Phi_k}{\mathcal{D}_k^{5/2}} + \hat{u}_2 \frac{15}{8} \sum_k \frac{P_{k,k}}{\mathcal{D}_k^{7/2}}$$
$$+ \hat{u}_3 \frac{15}{8} \sum_{k_1 k_2} \frac{Y_2(k_1,k_2)}{Y_0(k_1,k_2)^{7/2}} - \hat{u}_3 \frac{3}{2} \sum_{k_1 k_2} \frac{Y_3(k_1,k_2)}{Y_0(k_1,k_2)^{5/2}},$$
$$(13)$$

where $\hat{u}_L \equiv (2\pi)^{-3(L-1)/2} \bar{u}_L$ and $\hat{\Delta} \equiv (2\pi)^{-3/2} \Delta$. Note, however, that from the point of view of the Flory theory the spring term has an entropic origin and the proper conformational entropy should be defined as [18]

$$S_c = S - \frac{3\kappa N}{2} \mathcal{D}_{01}. \quad (14)$$

Here we should also introduce the monomer spatial correlations,

$$\mathcal{D}_{mm'} \equiv \overline{D_{mm'}}, \quad D_{mm'} = \frac{1}{3} \langle (\mathbf{X}_m - \mathbf{X}_{m'})^2 \rangle, \quad (15)$$

and their cumulants,

$$\overline{D_{mm'} D_{m''m'}}^{(c)} \equiv \overline{D_{mm'} D_{m''m'}} - \overline{D_{mm'}}\, \overline{D_{m''m'}}. \quad (16)$$

One can prove that these functions depend only on the differences of their indices, $k_1 = m - m'$ and $k_2 = m'' - m'$,

$$\mathcal{D}_{mm'} \equiv \mathcal{D}_k, \quad \overline{D_{mm'} D_{m''m'}}^{(c)} \equiv P_{k_1 k_2}. \quad (17)$$

Other definitions for Eq. (13) may be found in the Appendix.

To understand the phase behavior of the system we need to identify the important order parameters. First, the overall size of the chain is given by the average $\overline{R_g^2}$ of the squared radius of gyration,

$$R_g^2 = \sum_{q \neq 0} F_q. \quad (18)$$

Second, the glassy behavior is reflected in the cumulant of the squared radii of gyration [14]

$$\overline{R_g^2 R_g^{2(c)}} = \bar{\Delta}^{-2} \Upsilon^2, \quad \Upsilon = \sum_{q \neq 0} \varphi_{qq}. \quad (19)$$

And third, the phase-separation order parameter is

$$\Psi = \frac{1}{6N^2} \sum_{mm'} \overline{(\Lambda_m + \Lambda_{m'} - 2\lambda_0) D_{mm'}} = \sum_{q \neq p,\, q,p \neq 0} \varphi_{qp}. \quad (20)$$

Note that for just two types of monomers "A" and "B" with equal concentrations $n_A = n_B = \frac{1}{2}$ the latter reduces simply to

$$\Psi = [R_g^2(B) - R_g^2(A)]/2. \quad (21)$$

III. PHASE DIAGRAM OF THE MODEL

Here we present results of the numerical analysis of the equilibrium limit of Eqs. (10) and (11) obtained by setting the time derivatives to zero. Moreover, we require the expressions (12) and (13) for the free energy itself.

Traditionally [19–21] we work with the following combinations $\mathcal{L}=(k_B T/\kappa)^{1/2}$ and $\mathcal{T}=\zeta_b/\kappa$ as the units of size and time in the system. In the following, we have used the following particular choice of parameters: $k_B T=1$, $\kappa=1$, and $\zeta_b=1$, which fix $\mathcal{L}$ and $\mathcal{T}$ to be equal to unity.

The phase diagram is presented in Fig. 1. Phase (I) corresponds to the extended Flory coil. This normal homopolymerlike coil with increasing dispersion of disorder becomes what we call a random coil (III) after a rather soft continuous
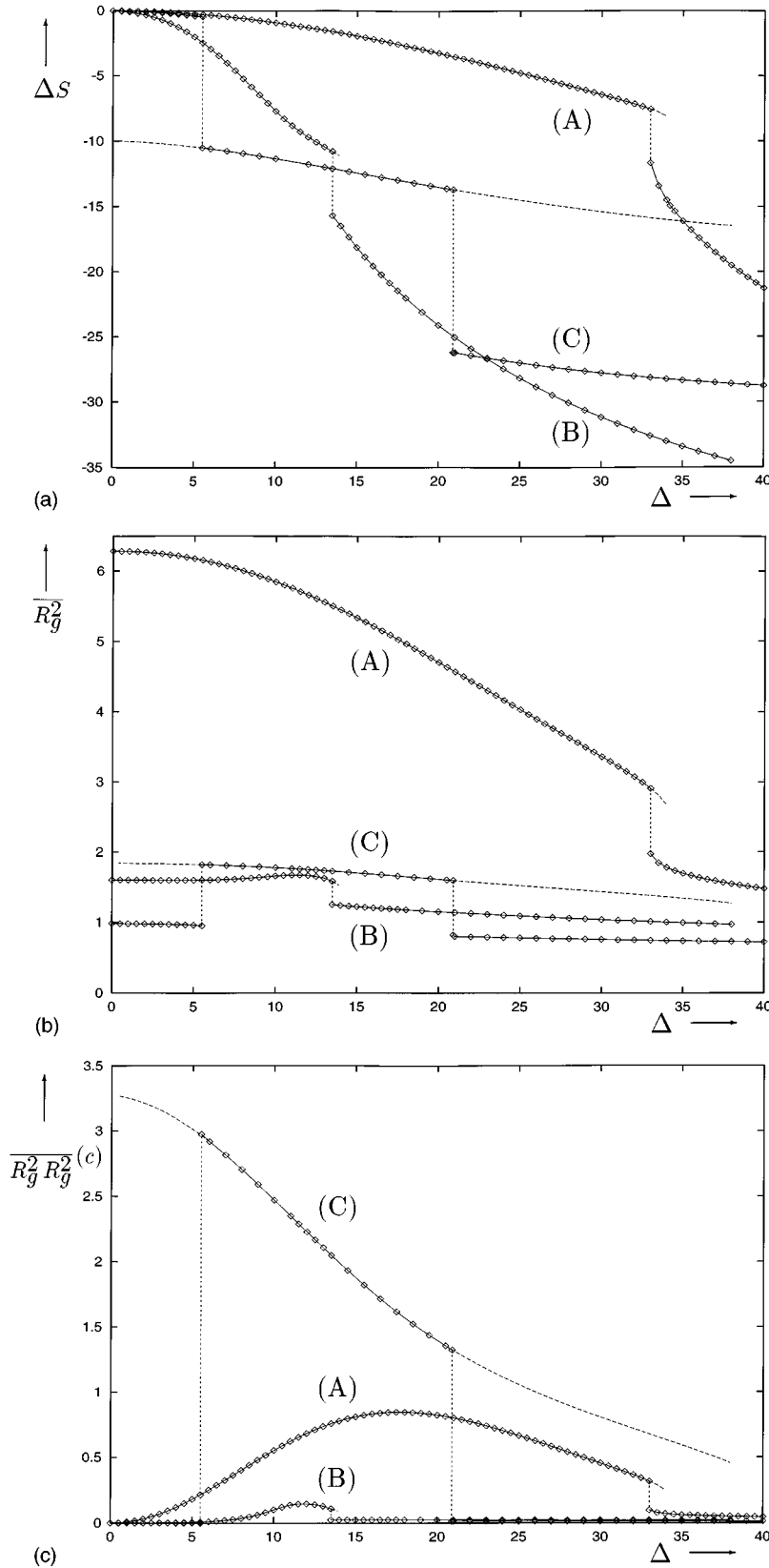
FIG. 2. Plots of observables vs the dispersion of disorder Δ (in units $k_B T \mathcal{L}^3$) for different values of $\overline{u}_2$ (in units $k_B T \mathcal{L}_3$): (a) $\overline{u^2} = 0$; (b) $\overline{u}_2 = -16$; and (c) $\overline{u}_2 = -35$. In (a) we present the conformational entropy change $\Delta S \equiv S_c(\Delta) - S_c(0)$ (in units $k_B T$); in (b), the mean-squared radius of gyration $\overline{R_g^2}$ (in units $\mathcal{L}^2$); and in (c), the glass order parameter $\overline{R_g^2 R_g^{2}}^{(c)}$ (in units $\mathcal{L}^4$).

transition. On passing this transition the squared radius of gyration $\overline{R_g^2}$ decreases somewhat while the ''glassy'' order parameter $\overline{R_g^2 R_g^{2(c)}}$ increases significantly. We believe that phase (III) is composed of relatively open coils with numerous loops. Indeed, in this region of the phase diagram the average two-body interaction is repulsive, but there are some strongly attractive units in the chain that can bind together forming loops. It is quite possible that this only leads to a renormalization of the Kuhn length and does not affect the fractal dimension of the coil. However, at present our

TABLE I. Values of the dispersion of disorder at the spinodal Δ_{sp}, and the critical curves Δ_{tr}, respectively, for the folding transition.

N	20	30	40	50	60
Δ_{sp}, $\bar{u}_2 = 0$	28.9	34.2	45.8	>400	>600
Δ_{tr}, $\bar{u}_2 = 0$	28.9	31.9	34.5	36.8	39.1
Δ_{sp}, $\bar{u}_2 = -25$	17.82	25.96	35.44	77.43	>600
Δ_{tr}, $\bar{u}_2 = -25$	15.7	17.0	18.3	19.5	20.6
Δ_{sp}, $\bar{u}_2 = -50$	41.5	59.2	79.1	196	>600
Δ_{tr}, $\bar{u}_2 = -50$	26.4	27.4	28.5	29.4	30.4

current numerical procedure was limited to relatively short polymers and did not allow us to resolve this question. This is a very interesting problem and we hope to return to its study at a later stage.

The collapse transition (the curve separating I and III from all other phases) is second order in the whole range of Δ. Beginning from the homopolymerlike coil (I) for small dispersions of disorder it leads to a homopolymerlike globule (II). That phase is often referred to as a liquidlike globule in the literature because the connectivity constraints there are manifested only at short distances along the chain.

In the next series of figures we draw the behavior of the conformational entropy [Fig. 2(a)], the squared radius of gyration [Fig. 2(b)], and the glass order parameter [Fig. 2(c)] vs the dispersion of disorder Δ for three different values of the second virial coefficient.

With increasing Δ the globule (II) undergoes a freezing transition. As evident from Fig. 2 the frozen globule (IV) has a smaller entropy, a larger size, and a pronounced glass order parameter. The freezing transition is continuous above the tricritical point and becomes first-order-like below it. In Fig. 1 we show the spinodal curve beyond which the homopolymerlike globule solution ceases to exist. Note that phase (IV) continues to exist to the left of that curve. To find the freezing transition curve we have to compare values of the free energy of the two solutions. It turns out that these values differ very slightly by a contribution proportional to Δ^4. As we have discussed in the Introduction we may not fully rely on the predictions of the method in such a case. Indeed, the free energy of the frozen globule solution remains somewhat lower than that of the normal globule for any small Δ—a conclusion that cannot be considered as physically reasonable. Thus although we may rely on the result for the spinodal curve, we are unable to reliably determine the freezing transition curve at the current order of the weak disorder expansion [10]. The resolution of this deficiency of the current version of the method lies in the inclusion of higher-order corrections. To do so is in principle a substantial technical task. Nevertheless, in Sec. IV we show that there is an unambiguous way to resolve the problem, albeit phenomenologically, within a simplified treatment discussed below.

We have no such difficulty for the first-order-like folding transition. Here the free energies on the two branches differ significantly (in Δ^2 order) and hence the folding transition curve is depicted along with the spinodal in Fig. 1. According to Fig. 2 the folded state (V) is characterized by very low entropy, compact size, smaller glass order parameter, and optimal phase separation of hydrophobic and hydrophilic units (see Figs. 2 and 7 in Ref. [10]). We believe that the

frozen globule is akin to the *molten* globule and the folded globule to the *native* state in proteins, respectively. This conjecture is somewhat justified also by the kinetics of the folding process discussed in Ref. [10].

An interesting observation here is that there is a pronounced region of the metastable frozen globule. The kinetic evolution after a rapid quench to the region bounded by the transition and spinodal curves will remain trapped in a frozen misfolded state for a long time related to the barrier height. This is in agreement with observation in numerous Monte Carlo simulations [22–26] that there is poor kinetic accessibility of the folded conformation for a generic class of sequences.

The situation deteriorates dramatically with increasing degree of polymerization according to Table I. Thus for sufficiently long chains the spinodal curve cannot be reached until the dispersion of disorder value Δ_{sp}, which tends to infinity exponentially quickly. This means that in our model, having all possible sequences of monomers characterized by a Gaussian distribution, the kinetic accessibility is very poor indeed. This is by no means surprising—the Gaussian distribution is too wide—as has been pointed out by us [10] and, earlier, by many others [27–29]. The way to make folding more efficient is to optimize the distribution by restricting acceptable sequences to a narrow subclass of sequences possessing good folding properties.

An intriguing point however is that, although the kinetic accessibility is impeded for long polymers, the folded state still remains the main free energy minimum for systems with quite small dispersions of disorder (see Table I). The more nontrivial issue of whether this remains true in the thermodynamic limit ($N \to \infty$) is answered in the affirmative in Sec. IV.

IV. LIMIT OF THE DENSE GLOBULE

In this section we analyze the limit of the high-density globule (i.e., our further considerations are valid only in the region $\hat{u}_2 < 0$), so that $\rho^{2/3} \sim (|\hat{u}_2|/\hat{u}_3)^{2/3} \gg \kappa$, and hence one may neglect the spring term. We start by noting that for a homopolymer in this limit the equations possess only a constant solution $\mathcal{F}_q \equiv \mathcal{F} = \text{const}$, and $\mathcal{D}_m \equiv \mathcal{D} = 2N\mathcal{F} = (\frac{4}{3})(2\hat{u}_3 N/|\hat{u}_2|)^{2/3}$ for the conformational modes q, $m \neq 0$. This is pretty much true for a normal liquidlike globule, since for sufficiently large m the function $\mathcal{D}_m$ quickly saturates to a constant. The latter behavior is simply the mathematical expression of the physical observation that the connectivity contributions of polymer chains are screened in the dense limit.

For a random copolymer we shall seek an analogous constant solution by requiring, in addition, that

$$\varphi_{q \neq p} \equiv \varphi = \text{const}, \quad \varphi_{qq} \equiv \widetilde{\varphi} = \text{const}. \quad (22)$$

It is now natural to introduce the rescaled variables,

$$r^2 = \frac{\mathcal{D}}{N^{2/3}}, \quad \chi = \frac{N^2 \varphi}{\Delta \mathcal{D}}, \quad w = \frac{2N^{3/2} \widetilde{\varphi}}{\Delta \mathcal{D}}. \quad (23)$$

The first definition obviously reflects the scaling of the compact globule size on the polymer length N. The second variable could be also understood as a dimensionless degree of the phase separation. Indeed, from the definition $\Psi = N^2 \varphi$, for the case of the binary distribution and equal concentration, by using Eq. (21), we obtain

$$\chi = \frac{1}{2} \frac{R_g^2(B) - R_g^2(A)}{R_g^2(B) + R_g^2(A)} \leq \frac{1}{2}. \quad (24)$$

As for the third variable, in that case one can show similarly that,

$$w^2 = \frac{\overline{R_g^2 R_g^{2(c)}}}{(\overline{R_g^2})^2} \leq 1. \quad (25)$$

We believe that Eq. (23) may represent an important generalization of scaling variables to the realm of random copolymers. Let us write out the specific energy ϵ, entropy s, and free energy a,

$$\epsilon = \frac{\mathcal{E}}{N}, \quad s = \frac{\mathcal{S}}{N}, \quad a = \frac{\mathcal{A}}{N}, \quad (26)$$

as well as the mobility per monomer ζ_b, and the characteristic time scale, τ,

$$\zeta_b = \frac{\zeta}{N}, \quad \tau = \frac{t}{N^{2/3}}. \quad (27)$$

Then, the specific energy (13) and entropy (12) reduce to the expressions,

$$\epsilon = r^{-3} \left[U_2[\chi, w] - \frac{3}{2} \hat{\Delta} \left(\chi + \frac{w}{\sqrt{N}} \right) \right] + r^{-6} \left(\frac{4}{3} \right)^{3/2} U_3[\chi, w], \quad (28)$$

$$s = \frac{3}{4} k_B [4 \log r + \log(1 - 4\chi^2 - w^2)]. \quad (29)$$

$$U_2[\chi, w] = \hat{u}_2 [1 + \tfrac{15}{8}(\chi^2 + w^2) + \cdots], \quad (30)$$

$$U_3[\chi, w] = \hat{u}_3 [1 + \tfrac{9}{2}(\chi^2 + w^2) + \cdots]. \quad (31)$$

The kinetic equations following from Eqs. (10) and (11) take form,

$$\frac{\zeta_b}{2} \frac{dr}{d\tau} = \frac{k_B T}{r} - \frac{1}{3} \frac{\partial \epsilon}{\partial r} = -\frac{1}{3} \frac{\partial a}{\partial r}, \quad (32)$$

$$\frac{\zeta_b}{2} \frac{d\chi}{d\tau} = -\frac{\chi}{r^2} \left(2k_B T + \frac{1 - 4\chi^2}{3\chi} \frac{\partial \epsilon}{\partial \chi} - \frac{4w}{3} \frac{\partial \epsilon}{\partial w} \right)$$
$$= -\frac{1 - 4\chi^2}{3r^2} \frac{\partial a}{\partial \chi} + \frac{4\chi w}{3r^2} \frac{\partial a}{\partial w}, \quad (33)$$

$$\frac{\zeta_b}{2} \frac{dw}{d\tau} = -\frac{w}{r^2} \left(2k_B T + \frac{4(1 - w^2)}{3w} \frac{\partial \epsilon}{\partial w} - \frac{4\chi}{3} \frac{\partial \epsilon}{\partial \chi} \right)$$
$$= -\frac{4(1 - w^2)}{3r^2} \frac{\partial a}{\partial w} + \frac{4\chi w}{3r^2} \frac{\partial a}{\partial \chi}, \quad (34)$$

Let us emphasize that the entropy in the ansatz (29), although it coincides with the approximate expression (12) restricted to constant variables in lower orders of expansion, is in fact exact within the limits of the Gaussian method. Thus kinetic equations (32) and (33) can be exactly obtained by differentiation of the free energy, Eqs. (29) and (28).

It is encouraging that the time derivative of the free energy is nonpositive due to the bounds (24) and (25) and the relation,

$$\zeta_b \frac{da}{d\tau} = -\frac{2}{3} \frac{\partial a}{\partial \mathbf{X}} \cdot M \cdot \frac{\partial a}{\partial \mathbf{X}}, \quad (35)$$

where we have used the vector notations $\mathbf{X} = (r, \chi, w)$ and M is a matrix with a positive definite determinant

$$\det M = \frac{4}{r^4}(1 - 4\chi^2 - w^2) > 0. \quad (36)$$

The validity of similar approximate treatments was discussed by us in Ref. [30]. Direct comparison with the exact results here confirms that the approximation is fairly good for sufficiently dense globule and large N. In Figs. 3 and 4 we support this assertion by presenting a part of the phase diagram and dependences of the variables r, w, and χ vs Δ. We observe here the same qualitative behavior as in the complete scheme of Sec. III. Really, for phase (IV) the glass parameter w is very large and the size r is larger there than for phase (II). For phase (V) the phase separation χ is larger, while r and w are smaller than for (IV). In this simplified treatment it is trivial to take the thermodynamic limit $N \to \infty$. Simple numerical analysis of Eqs. (32)–(34), (29) and (28) shows that the phase boundary of the folded state is weakly dependent on N and that this state remains the main free energy minimum for phase (V) even in the thermodynamic limit.

As a criticism of the approximation for finite values of N and κ we note that the constant ansatz underestimates the conformational entropy of the frozen globule (IV). In Ref. [10] we alluded to an observation that such a globule consists of locally frozen clusters along the chain, and therefore is sensitive to the spring term. As a result the frozen phase (IV) is somewhat narrower in the approximate scheme compared to that of Fig. 1. For similar reasons we cannot really trust the N dependence of the folding spinodal line here and should use, instead, data from Table I. In other respects the current approximate scheme presents an attractive and easily tractable set of only three equations for large-scale order parameters that still correctly describe the essential conforma-
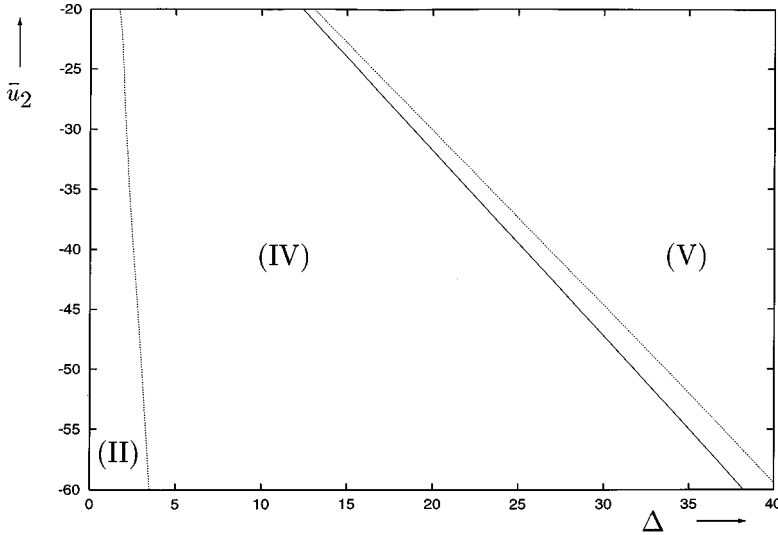
FIG. 3. Part of the phase diagram for the dense globule in terms of the second virial coefficient $\bar{u}_2$ (in units $k_B T \mathcal{L}^3$), and the dispersion of disorder Δ (in units $k_B T \mathcal{L}^3$), obtained from the reduced self-consistent equations (32)–(34). Solid lines represent first-order-like transitions, dashed lines represent continuous transitions, and dotted lines represent ''spinodal'' curves. Here all parameters have the same values as in Fig. 1, but $\kappa = 0$.

tional transitions of the dense globule. It appears to be a reasonable generalization of simple Flory-type theories to random copolymers, and has the additional merit of describing the kinetics of conformational changes of the globule.

Despite certain improvements of the entropy (29) there still remains the problem with determination of the freezing transition curve. However, now we may attempt to address it phenomenologically along similar lines of thinking. Indeed, let us improve the theory by adding the higher-order terms in Eq. (30),

$$U_2 = \hat{u}_2 (1 + \alpha_2 y + \alpha_4 y^2 + \cdots), \tag{37}$$

where $y \equiv \chi^2 + w^2$ and $\alpha_2 = \frac{15}{8}$. We must, strictly speaking, evaluate the higher-order terms renormalizing Δ and u_3 and so on, but let us assume for the moment that they are less significant for the problem at hand. The coefficient α_4 can be found by the requirement that the folding transition curve behaves as $\Delta \to 0$ for $\hat{u}_2 \to -\infty$. This is equivalent to requiring that for sufficiently large $|\hat{u}_2|$ we have

$$(\epsilon_{lg} - \epsilon_{gl})(\Delta = 0) \sim -\frac{|\hat{u}_2|}{\hat{u}_3} \left(1 - \frac{[1 + (\frac{15}{8})y + \alpha_4 y^2]^2}{1 + (\frac{9}{2})y} \right) \leqslant 0. \tag{38}$$

By setting the discriminant of the resulting cubic equation to zero we obtain $\alpha_4 = -\frac{9}{16}$. For this choice of α_4 it is possible, therefore, to cure the theory and determine the freezing transition law and we find $\Delta_{\text{freez}} \sim |\hat{u}_2|^{-\alpha}$, where $\alpha = 0.96 \pm 0.04$. We have derived this result phenomenologically, but it is consistent with what one would expect for the freezing transition $\Delta_{\text{freez}} \sim \rho^{-1} \sim |\hat{u}_2|^{-1}$ obtained from the replica formalism in Refs. [31] and [15]. Somewhat different expressions have been derived for the freezing transition in the symmetric random ''charge'' [16] model [32,33].

V. CONCLUSION

We regard this work as the conclusion to our paper, Ref. [10], in which we proposed a self-consistent method for the model of a Gaussian random amphiphilic copolymer. There
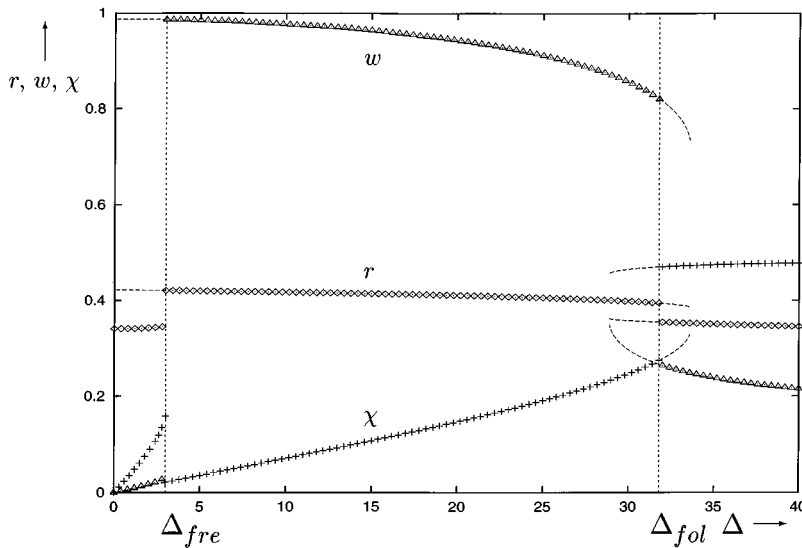

FIG. 4. Plots of the dimensionless quantities r, w, and χ vs the dispersion of disorder Δ (in units $k_B T \mathcal{L}^3$) for $\bar{u}_2 = -50$ (in units $k_B T \mathcal{L}^3$) and other parameters as in Fig. 3. Points correspond to the values of observables in the global minimum of the free energy, dashed lines correspond to values in metastable minima; vertical dotted lines correspond to the points of discontinuous transitions.

we discussed the kinetics of the folding process and here we investigated the equilibrium diagram of states of the model.

In many respects we found the method satisfactory in attacking such a hard problem. Unfortunately, it has a deficiency in describing the freezing transition curve. However, it can be fixed by phenomenologically improving the expression for the internal energy and self-consistently determining the coefficient of the higher-order correction. We believe that a similar coarse-grained statistical description provides the fundamental methodology that in future will permit further progress in building more detailed models of proteins.

One of the principal conclusions of our consideration is that the Gaussian distribution of sequences is too wide to possess any good folding properties in the average. Despite the thermodynamic stability of the folded state for long chains, its kinetic accessibility is very impeded. Thus to achieve the folded conformation a polymer should generally overcome a potential barrier whose height grows with the chain length. Selection and design of good folding sequences, therefore, remains the main issue in the fundamental problem of protein folding. While we presently study rather simple models of random copolymers, we believe that the direction we have laid out will have increasing relevance for biological problems.

ACKNOWLEDGMENTS

APPENDIX: SOME DEFINITIONS

For a ring polymer of length N the Fourier transformation is defined as

$$\mathbf{X}_m = \sum_{q=0}^{N-1} f_m^{(-q)} \mathbf{x}_q, \quad \mathbf{x}_q = \frac{1}{N} \sum_{m=0}^{N-1} f_m^{(q)} \mathbf{X}_m, \quad (A1)$$

$$f_m^{(q)} \equiv \exp\left(\frac{2\pi i q m}{N}\right). \quad (A2)$$

The two-point equal-time correlation functions (15) may be written in terms of the Fourier modes as

$$D_{mm'} = \sum_q d_{mm'}^{(q)} F_q, \quad (A3)$$

$$d_{mm'}^{(q)} = 2\left(1 - \cos\frac{2\pi q(m-m')}{N}\right). \quad (A4)$$

Analogously we deduce the three-point functions

$$D_{mm'm''} \equiv \tfrac{1}{3}\langle (\mathbf{X}_m - \mathbf{X}_{m'})(\mathbf{X}_{m''} - \mathbf{X}_{m'}) \rangle = \sum_q d_{mm'm''}^{(q)} F_q, \quad (A5)$$

$$d_{mm'm''}^{(q)} = \tfrac{1}{2}(d_{mm'}^{(q)} + d_{m''m'}^{(q)} - d_{mm''}^{(q)}). \quad (A6)$$

In Eq. (13) we have used the following set of definitions:

$$Y_0(k_1,k_2) = \mathcal{D}_{k_1}\mathcal{D}_{k_2} - \mathcal{D}_{k_1 k_2}^2, \quad (A7)$$

$$Y_2(k_1,k_2) = \mathcal{D}_{k_1}^2 P_{k_2,k_2} + \mathcal{D}_{k_2}^2 P_{k_1,k_1} + 4\mathcal{D}_{k_1 k_2}^2 P_{k_1 k_2,k_1 k_2}$$
$$+ 2\mathcal{D}_{k_1}\mathcal{D}_{k_2} P_{k_1,k_2} - 4\mathcal{D}_{k_1 k_2}(\mathcal{D}_{k_2} P_{k_1,k_1 k_2}$$
$$+ \mathcal{D}_{k_1} P_{k_2,k_1 k_2}), \quad (A8)$$

$$Y_3(k_1,k_2) = P_{k_1,k_2} - P_{k_1 k_2,k_1 k_2}, \quad (A9)$$

$$\Phi_k = \sum_{qp} d_k^{(q,p)} \varphi_{gp}. \quad (A10)$$

We have also denoted

$$P_k^{(s)} = \sum_p d_k^{(p,p+s)} \varphi_{p,p+s}, \quad (A11)$$

$$P_{k_1,k_2} = \overline{\Delta}^{-2} \sum_s P_{k_1}^{(s)} P_{k_2}^{(s)} = \overline{D_{k_1} D_{k_2}}^{(c)} \quad (A12)$$

with the coefficients

$$d_k^{(q,p)} = \tfrac{1}{2}(d_k^{(q)} + d_k^{(p)} - d_k^{(q-p)}). \quad (A13)$$

[1] T. Garel and H. Orland, Europhys. Lett. **6**, 307 (1988); **6**, 597 (1988); (unpublished).

[2] G. Taubes, Science **271**, 1493 (1996); R. A. Bessen *et al.*, Nature (London) **375**, 698 (1995).

[3] G. H. Fredrickson, S. T. Milner, and L. Leibler, Macromolecules **25**, 6341 (1992).

[4] H. Frauenfelder, in *Structure and Dynamics of Nucleic Acids Proteins and Membranes*, edited by E. Clementi and S. Chin (Plenum, New York, 1986).

[5] P. G. Wolynes, in *Spin Glass Ideas in Biology*, edited by D. Stein (World Scientific, Singapore, 1991).

[6] T. E. Creighton, *Protein Folding* (Wiley, New York, 1992).

[7] R. Elber, *New Developments in Theoretical Studies of Proteins* (World Scientific, Singapore, 1994).

[8] R. L. Baldwin, Nature (London) **346**, 409 (1990).

[9] A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).

[10] E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson, Phys. Rev. E **54**, 4071 (1996).

[11] S. E. Radford, C. M. Dobson, and P. E. Evans, Nature (London) **358**, 302 (1992).

[12] D. Thirumalai and S. A. Woodson, Acc. Chem. Res. **29**, 433 (1996).

[13] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[14] M. Mezard and G. Parisi, J. Phys. (France) I **1**, 809 (1991).

[15] A. Moskalenko, Yu. A. Kuznetsov, and K. A. Dawson, J. Phys. (France) II **7**, 409 (1997).

[16] We note that the widely used quadratic term in the disorder

variables corresponds to the Edwards free energy functional constructed as the virial expansion, $\Sigma_L \int d\mathbf{y}(\rho_p(\mathbf{y}))^L$, in terms of the pseudodensity, $\rho_p(\mathbf{y}) = \Sigma_m \Lambda_m \delta(\mathbf{X}_m - \mathbf{y})$. This model was exploited by Fredrickson, Milner, and Leibler [3] and many others, and although it is clearly suitable for a mean-field Flory-type theory, microscopically it corresponds to rather nonlocal monomer-solvent interactions. Thus the model with the linear term in Λ_n is adequate for studying the hydrophobic interaction of monomers that differ in their interaction properties with the solvent molecules but not with each other. On the other hand, the model with the quadratic term in disorder variables, often called the random ''charge'' model, is more appropriate for describing either true charges or nonlocal effective monomer-solvent interactions arising after coarse-graining of models with complex intramolecular potentials.

[17] These kinetic equations describe the nonequilibrium time evolution of the state of the system rather than its autocorrelation dynamic properties in the glassy phase. For the latter, a well-known approach is based on the Sompolinsky and Zippelius treatment [34] that uses the path-integral representation of the transition probability. There are some intrinsic difficulties in their approach related to the choice of the initial conditions. These difficulties do not appear here since we consider kinetics after a quench from the ergodic homopolymer state.

[18] F. Ganazzoli, R. La Ferla, and G. Allegra, Macromolecules **28**, 5285 (1995).

[19] E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson, J. Chem. Phys. **102**, 1816 (1995).

[20] Yu. A. Kuznetsov, E. G. Timoshenko, and K. A. Dawson, J. Chem. Phys. **104**, 3338 (1996).

[21] E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson, Phys. Rev. E **53**, 3886 (1996).

[22] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci.

U.S.A. **84**, 7524 (1987); P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).

[23] K. F. Lau and K. A. Dill, Macromolecules **22**, 3986 (1989); H. S. Chan and K. A. Dill, *ibid.* **22**, 4559 (1989); K. A. Dill, K. M. Fiebig, and H. S. Chan, Proc. Natl. Acad. Sci. U.S.A. **90**, 1942 (1993).

[24] C. J. Camacho and D. Thirumalai, Phys. Rev. Lett. **71**, 2505 (1993); Proc. Natl. Acad. Sci. U.S.A. **90**, 6369 (1993); **90**, 1277 (1995).

[25] J. D. Honeycutt and D. Thirumalai, Biopolymers **32**, 695 (1992); D. Thirumalai, J. Phys. (France) I **5**, 1457 (1995).

[26] P. E. Leopold, M. Montal, and J. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **89**, 9721 (1992); N. D. Socci and J. N. Onuchic, J. Chem. Phys. **103**, 4732 (1995).

[27] K. Yue and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **89**, 4163 (1992).

[28] E. I. Shakhnovich and A. M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1993).

[29] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, J. Chem. Phys. **101**, 8246 (1994); Macromolecules **28**, 2218 (1995); J. Chem. Phys. **103**, 9482 (1995).

[30] E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson, J. Stat. Phys. (to be published).

[31] A. Moskalenko and K. A. Dawson, J. Chem. Phys. **103**, 9886 (1995).

[32] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Phys. Rev. E **51**, 3381 (1995); J. Chem. Phys. **101**, 8246 (1994).

[33] E. I. Shakhnovich and A. M. Gutin, J. Phys. A **22**, 1647 (1989); J. Chem. Phys. **93**, 5967 (1990); C. D. Sfatos, A. M. Gutin, and E. I. Shakhnovich, Phys. Rev. E **48**, 465 (1993).

[34] H. Sompolinsky and A. Zippelius, Phys. Rev. Lett. **45**, 359 (1981); Phys. Rev. B **25**, 274 (1982); A. Houghton, S. Jain, and A. P. Young, *ibid.* **28**, 2630 (1983).