PHYSICAL REVIEW E **66**, 031803 (2002)

Bethe approximation for a hydrophobic-polar random copolymer

M. Pretti
Istituto Nazionale per la Fisica della Materia (INFM) and Dipartimento di Fisica, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy

(Received 27 July 2001; revised manuscript received 21 March 2002; published 23 September 2002)

A previous analysis of the configurational statistics of a lattice self-avoiding walk, based on the pair approximation of the cluster variation method (Bethe approximation), is generalized to describe a random copolymer in dilute solution, in which monomers interact with one another and with the solvent. The phase diagram is worked out numerically and the coil-globule transition is obtained. Entropy and internal energy in the coil phase, as well as the Θ point are evaluated analytically, by means of a Landau expansion of the variational free energy. A detailed analysis is carried out for the case in which the only interaction present is an attraction between monomers of one species. Such a model has been proposed in the literature to mimic a random copolymer made up of hydrophobic and polar monomers in water solution, a physical system which is believed to be relevant to understanding the behavior of proteins.

DOI: 10.1103/PhysRevE.66.031803 PACS number(s): 82.35.Jk, 05.50.+q, 05.70.Fh, 64.60.Cn

I. INTRODUCTION

The Bethe approximation [1] is a common statistical mechanical technique to study lattice systems. It improves upon the simple mean-field theory, by taking into account correlations between nearest neighbor (NN) sites, and can be reformulated with different points of view. By a simplified counting of lattice states, it has been presented as a quasichemical approximation [2]. By a suitable truncation of the entropy cumulant expansion it can be seen as the lowest step (pair approximation [3,4]) of a hierarchy of approximations that takes into account correlations up to arbitrarily large clusters, and is known as the cluster variation method (CVM) [5]. The pair approximation was first applied to lattice polymer statistics by Aguilera-Granja and Kikuchi, who treated the problem of (polydisperse) polymer solutions [6–8], and obtained several improvements over the classical Flory theory [9,10]. More recently the pair approximation was generalized to treat the configurational statistics of a single self-avoiding walk (SAW) in the grand canonical ensemble [11], which is the usual way to model a single linear polymer in solution [12,13]. In Ref. [11] the authors considered a semiflexible polymer model, and noticed several improvements over the corresponding mean-field theory as well, in spite of the simplicity of the pair approximation. In this paper we have the purpose of extending this kind of approach to the case of a random copolymer in which monomers of the two species interact with one another and with the solvent, introducing suitable approximations to treat the quenched disorder of the monomer sequence.

Copolymer models have attracted great attention in the latest years, especially in the case of monomer species distinguished on the basis of their degree of hydrophilicity (hydrophobic-hydrophilic copolymers), because of connections to the protein folding problem. The importance of hydrophobicity to protein folding was shown by Dill [14], who suggested that the main driving force of the process was the tendency of polar (hydrophilic) amino acid monomers to shield nonpolar (hydrophobic) monomers from the water environment. Of course the formation of a peculiar native structure needs also the presence of a particular sequence of amino acids and a randomly chosen sequence usually does not fold. Nevertheless, studying a particular case of the simple model treated in this paper (by a mean-field-like approach on a ‘‘spherical’’ lattice), Dill obtained important qualitative information about the folding process [15] and the thermal stability [16] of globular proteins. Since then, the so-called H-P (hydrophobic-polar) model has been extensively investigated from different points of view. Exact enumeration studies on finite (fixed) H-P sequences have been devoted to find the sequences that are likely to be good folders [17] or to determine the designability of native states [18]. More analytical investigations, based on the replica theory, have been devoted to determine general properties of the random H-P copolymer model, mainly in the presence of Gaussian disorder [19,20]. In Ref. [19] the authors found that in the case of weak hydrophobicity the ordinary Θ collapse is replaced by an unusual first order collapse and that the entropy of the corresponding compact state is strongly reduced with respect to the ordinary collapsed phase. Moreover, in Ref. [20] a replica symmetry breaking has been observed, corresponding to a frozen state, in addition to the ordinary (molten) compact state.

In this paper we shall mainly focus on the Θ transition and related properties, especially in comparison with the case of a purely H homopolymer. As previously mentioned, the model is quite general in principle, and includes both monomer-monomer and monomer-solvent interactions. Nonetheless, as observed by Dill [15], it is possible to model a H-P copolymer by just retaining an attractive interaction energy between H monomers, neglecting all other interactions. In the framework of the Bethe approximation, we shall present detailed numerical results for this simpler case, although the analytical treatment is general. Due to the fact that pair correlations are taken into account exactly, we observe a more ‘‘realistic’’ behavior of the model, which could not be reproduced by a mean-field theory. Some features, such as a monomer rearrangement with H-H preference, are shown also in the coil phase, whose properties are determined analytically, by means of a Landau expansion. The

1063-651X/2002/66(3)/031803(11)/$20.00

same expansion allows us to determine analytically the Θ point, at which the second virial coefficient vanishes. It is then possible to analyze easily how the Θ-point location is affected by the percentage of H monomers in the chain, and in principle also by the interaction parameters and by the lattice coordination number.

We now give an outline of the paper. In Sec. II we describe the model in some more detail and recall in short a general way of treating thermodynamic systems with quenched disorder in the grand canonical ensemble. In Sec. III we derive the Bethe approximation of the grand potential. The equations used to determine its minima with the required constraints, and the most technical details of the numerical method are left to an appendix. In Sec. IV we present numerical results for the special case mentioned above (H-H interaction only). In Sec. V we perform the Landau expansion of the grand potential in powers of the density, and analytically determine the Θ transition, which is then investigated as a function of the relative monomer concentration. Finally, Sec. VI is devoted to an overall discussion, to a comparison with previous results, and to some concluding remarks.

II. THE MODEL AND THE FREE ENERGY

We consider a SAW on a lattice. Each site visited by the walk represents a monomer. Two different kinds of monomers are placed at random along the chain, that is, we assign a random variable $s=\pm 1$ to each monomer. In the following $s=+1$ will denote a H monomer and $s=-1$ will denote a P monomer. The probability distributions (PDs) of the random variables are fixed (quenched variables). Two monomers of species s, s' that are nonconsecutive along the chain and come in contact, i.e., are placed on two NN sites, are assumed to interact with an energy $\epsilon_{s,s'}$. Empty sites represent clusters of solvent (water) molecules, and a monomer-solvent interaction energy γ_s, depending on the monomer species s, is assigned to each monomer-solvent contact. Such a model can be readily mapped onto a model with monomer-monomer interactions only, but for clarity we shall treat all interactions explicitly in the following, except in Sec. V. We shall assume that the binary random variables that determine the sequence of H-P monomers are statistically independent with equal distributions: P_s will denote the probability that a monomer is of species s.

The randomness of the H-P sequence is a form of quenched disorder: in the remainder of this section we briefly summarize the way we shall deal with this problem. We first introduce the canonical free energy for a generic random system, along the lines of the procedure proposed by Morita [21] (see also Ref. [22]), giving then a generalization to the grand canonical ensemble, which is more convenient for polymer lattice models [13]. Let us consider a finite system at fixed volume and particle number N (we shall not denote explicitly the dependence on N in the canonical ensemble). Suppose that h is a set of random variables representing the quenched configuration of the system (in our case the H-P sequence), whereas x is another set of random variables representing the annealed degrees of freedom of the systems (that is, for instance, the positions of the monomers). The Helmholtz free energy associated with a fixed N-monomer sequence h can be written in terms of the probability distribution of x conditioned on h, denoted by $\mathcal{P}(x|h)$, in the following way:

$$F(h)=\sum_x \mathcal{P}(x|h)[\mathcal{H}(x,h)+k_BT\ln\mathcal{P}(x|h)], \tag{1}$$

where $\mathcal{H}(x,h)$ is the Hamiltonian of the system, k_B is the Boltzmann constant, and T is the absolute temperature. $\mathcal{P}(h)$ being the PD of the quenched sequences h, the quenched free energy can be obtained by averaging over h as

$$F=\sum_h \mathcal{P}(h)F(h), \tag{2}$$

which is easily rewritten as

$$F=\sum_{x,h} \mathcal{P}(x,h)[\mathcal{H}(x,h)+k_BT\ln\mathcal{P}(x,h)]+T\widetilde{S}, \tag{3}$$

where $\mathcal{P}(x,h)$ is the joint PD of the annealed (x) and quenched (h) degrees of freedom for the N-particle system, and

$$\widetilde{S}=-k_B\sum_h \mathcal{P}(h)\ln\mathcal{P}(h) \tag{4}$$

is the information entropy of the random (N-monomer) sequence, or more generally the entropy associated with the quenched disorder. A canonical variational principle holds, namely, the equilibrium PD $\mathcal{P}(x,h)$ is the one that minimizes F with the constraint

$$\sum_x \mathcal{P}(x,h)=\mathcal{P}(h). \tag{5}$$

In order to move to the grand canonical ensemble, one has to define the grand potential in the usual way as

$$\Omega=\sum_{N=0}^{\infty} \mathcal{P}(N)[F_N-\mu N+k_BT\ln\mathcal{P}(N)], \tag{6}$$

where μ is the chemical potential, $\mathcal{P}(N)$ is the probability that the system contains N particles, and F_N is the (quenched) free energy for the N-particle system. If the latter quantity is expressed via Eq. (3), in which all probabilities are conditioned to the particle number N and the N-particle Hamiltonian is denoted by $\mathcal{H}_N(x,h)$, we obtain

$$\Omega=\sum_{N=0}^{\infty}\sum_{x,h} \mathcal{P}(x,h,N)[\mathcal{H}_N(x,h)-\mu N+k_BT\ln\mathcal{P}(x,h,N)]$$
$$+T\sum_{N=0}^{\infty} \mathcal{P}(N)\widetilde{S}_N, \tag{7}$$

where $\mathcal{P}(x,h,N)$ is the joint PD of the particle number N and of the x and h degrees of freedom, and $\widetilde{S}_N$ is the information entropy of an N-monomer sequence, given by Eq. (4), in which the probabilities are conditioned on N. Also, a grand canonical variational principle holds: The equilibrium joint PD $\mathcal{P}(x,h,N)$ is determined as the one that minimizes Ω with the constraint

$$\sum_{x} \mathcal{P}(x,h,N)=\mathcal{P}(h,N), \tag{8}$$

equivalent to Eq. (5). Let us finally notice that in our model H and P monomers are placed at random along the chain with no correlation, that is, quenched variables are statistically independent (with equal distribution P_s). Hence the information entropy of an N-monomer sequence is

$$\tilde{S}_N=-Nk_B\sum_{s=\pm 1} P_s \ln P_s\,, \tag{9}$$

and the grand potential can thus be written as

$$\Omega=\sum_{N=0}^{\infty}\sum_{x,h} \mathcal{P}(x,h,N)[\mathcal{H}_N(x,h)-\tilde{\mu}N+k_BT\ln\mathcal{P}(x,h,N)], \tag{10}$$

where

$$\tilde{\mu}=\mu+k_BT\sum_{s=\pm 1} P_s \ln P_s\,. \tag{11}$$

In the next section we shall derive the Bethe approximation for the grand potential (10).

III. BETHE APPROXIMATION

Let us consider a generic lattice with $\mathcal{N}$ sites and coordination number q, and a polymer, i.e., a SAW, on the lattice. We see that defining the number N of monomers in the lattice, their positions x, and their chemical species h is equivalent to defining, for each lattice site, the following things: (i) whether the site is empty or visited by the walk and, in the latter case, (ii) the configuration of the monomer on that site, that is, the directions of the two chain segments toward the previous and next monomer, and (iii) the monomer species H or P. It is then useful to define, for each site, a state variable $k=-n,\ldots,\ n$, such that $k=0$ if the site is empty (occupied by solvent), $k>0$ $(k<0)$ if the site is occupied by a H (P) monomer, and $|k|=1,\ldots,\ n$ if the two chain segments (toward the previous and next monomer) choose one of the

$$n=\binom{q}{2} \tag{12}$$

possible configurations. Thus we have transformed a polymer lattice system into a Potts-like model in which, to preserve polymer connectivity and the self-avoiding condition, site state variables have to satisfy some constraints. In particular, if sites 1 and 2 are NNs, and the configurational state of site 1 is such that it is linked (not linked) to site 2, then also site 2 must be in a state linked (not linked) to site 1. The Bethe approximation, taking into account pair correlations in an exact way, is able to impose such constraints. The joint PD $\mathcal{P}(x,h,N)$, introduced in the previous section, is equivalently expressed by the joint PD $\mathcal{P}(k_1,\ldots,k_{\mathcal{N}})\equiv\mathcal{P}(k)$ of the state variables of all lattice sites. The grand potential of the system can then be written as

$$\Omega=\sum_{k} \mathcal{P}(k)[\mathcal{H}(k)-\tilde{\mu}N(k)+k_BT\ln\mathcal{P}(k)], \tag{13}$$

where $\mathcal{H}(k)$ is a lattice Hamiltonian, including NN pair energy contributions $\epsilon_{s,s'}$ and γ_s (as described in the previous section), and infinite energy penalties assigned to NN pair states that violate connectivity constraints.

According to the Bethe approximation, the lattice state PD $\mathcal{P}(k)$ can be approximated by

$$\mathcal{P}(k)\approx\prod_{\langle i,j\rangle} p_{k_i,k_j}^{(i,j)}\left[\prod_{i} p_{k_i}^{(i)}\right]^{-(q-1)}, \tag{14}$$

where $\Pi_{\langle i,j\rangle}$ denotes a product over all NN pairs, Π_i a product over all sites, $p_{k,k'}^{(i,j)}$ is a pair PD (i.e., the probability that the NN sites i,j are, respectively, in the states k, k'), and $p_k^{(i)}$ is a site PD (i.e., the probability that the site i is in the state k). Due to the fact that the interaction energy $\mathcal{H}(k)$ and the chemical potential term $-\tilde{\mu}N(k)$ can be written as a sum of terms depending only on NN pair states as

$$\mathcal{H}(k)-\tilde{\mu}N(k)=\sum_{\langle i,j\rangle} h_{k_i,k_j}^{(i,j)}, \tag{15}$$

where $h_{k_i,k_j}^{(i,j)}$ must be defined in a suitable way, Eq. (14) substituted into Eq. (13) yields

$$\beta\Omega\approx\sum_{\langle i,j\rangle}\sum_{k=-n}^{n}\sum_{k'=-n}^{n} p_{k,k'}^{(i,j)}(\beta h_{k,k'}^{(i,j)}+\ln p_{k,k'}^{(i,j)}) -(q-1)\sum_{i}\sum_{k=-n}^{n} p_k^{(i)}\ln p_k^{(i)}, \tag{16}$$

where as usual $\beta=1/k_BT$. The above expression of the grand potential can be further simplified. First one assumes homogeneity and isotropy, whence it is sufficient to evaluate the average contribution to the grand potential due to a fixed pair of NN sites. Secondly, it is possible to show that, in the framework of the Bethe approximation applied to a polymer with zero stiffness [11], each one of the n local chain configurations has the same probability (which, however, may be different for H and P monomers). With respect to a fixed pair of NN sites, say 1 and 2, it is then only necessary to distinguish configurations of site 1 that are linked (or not linked) to site 2 by a chain segment, and vice versa for site 2 with respect to site 1. We shall then consider for site 1 the state variable $k=-2,\ldots,2$ defined as follows: $k=0$ if the site is occupied by solvent, $k>0$ $(k<0)$ if the site is occupied by a H (P) monomer, $|k|=1$ if the monomer at site 1 is linked to site 2 by a chain segment (the two monomers are consecutive along the chain), and $|k|=2$ if the monomer is not linked to site 2 (site 2 is occupied by solvent or by a nonconsecutive

TABLE I. Site states, defined with respect to a fixed NN site (right column), and corresponding values of the state variable (left column). H,P monomers are denoted, respectively, by ⊕, ⊖; solvent (water) molecules by ⊙. The reference NN site is denoted by a dot. Oblique segments denote chemical bonds in generic directions except the reference one.

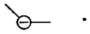
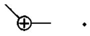
k	state
-2	
-1	
0	
$+1$	
$+2$	

TABLE II. Pair energies $h_{k,k'}$: $\epsilon_{s,s'}$ denote interaction energies between monomer species $s,s'=\pm$; γ_s denote interaction energies between a monomer species $s=\pm$ and the solvent; $\tilde{\mu}$ is defined by Eq. (11). $h_{k,k'}=\infty$ corresponds to pair states k,k' that violate connectivity constraints.

	k'				
k	-2	-1	0	$+1$	$+2$
-2	$\epsilon_{-,-}-2\dfrac{\tilde{\mu}}{q}$	∞	$\gamma_{-}-\dfrac{\tilde{\mu}}{q}$	∞	$\epsilon_{+,-}-2\dfrac{\tilde{\mu}}{q}$
-1	∞	$-2\dfrac{\tilde{\mu}}{q}$	∞	$-2\dfrac{\tilde{\mu}}{q}$	∞
0	$\gamma_{-}-\dfrac{\tilde{\mu}}{q}$	∞	0	∞	$\gamma_{+}-\dfrac{\tilde{\mu}}{q}$
$+1$	∞	$-2\dfrac{\tilde{\mu}}{q}$	∞	$-2\dfrac{\tilde{\mu}}{q}$	∞
$+2$	$\epsilon_{+,-}-2\dfrac{\tilde{\mu}}{q}$	∞	$\gamma_{+}-\dfrac{\tilde{\mu}}{q}$	∞	$\epsilon_{+,+}-2\dfrac{\tilde{\mu}}{q}$

monomer). The state of site 2 can be defined in an analogous way by a variable k': $|k'|=1$ ($|k'|=2$) if the monomer at site 2 is linked (not linked) to site 1. The reduced "encoding" of states is displayed in detail in Table I. The value of a site variable does not define an elementary state any longer, rather a number of elementary states, that is, a degenerate state. It is easy to verify that the degeneracy of k is

$$w_k=\binom{q-1}{|k|}. \tag{17}$$

Defining the grand potential per site as $\omega=\Omega/\mathcal{N}$ in the thermodynamic limit $\mathcal{N}\to\infty$, according to the above discussion we can finally write

$$\beta\omega=\frac{q}{2}\sum_{k=-2}^{2}\sum_{k'=-2}^{2} w_k w_{k'} p_{k,k'}^{(1,2)}(\beta h_{k,k'}+\ln p_{k,k'}^{(1,2)}) - \frac{q-1}{2}\sum_{i=1,2}\sum_{k=-2}^{2} w_k p_k^{(i)}\ln p_k^{(i)}, \tag{18}$$

where $q/2$ is the number of NN pairs per site, the pair energies $h_{k,k'}$ are those of Table II, and the site PDs $p_k^{(1)}$ and $p_{k'}^{(2)}$ are derived from the pair PD $p_{k,k'}^{(1,2)}$ as marginal distributions in the following way:

$$p_k^{(1)}=\sum_{k'=-2}^{2} w_{k'} p_{k,k'}^{(1,2)},$$

$$p_{k'}^{(2)}=\sum_{k=-2}^{2} w_k p_{k,k'}^{(1,2)}. \tag{19}$$

Of course, according to the homogeneity hypothesis, the last two distributions must be the same, that is,

$$p_k^{(1)}=p_k^{(2)} \tag{20}$$

(for all k), which is implied by the symmetry of the pair PD

$$p_{k,k'}^{(1,2)}=p_{k',k}^{(1,2)}. \tag{21}$$

We have thus written an approximate grand potential per site ω as a function of only the pair probabilities $p_{k,k'}^{(1,2)}$, the site probabilities depending on them via Eqs. (19). Let us notice that ω, as it has been derived, is actually the excess grand potential of the polymer-solvent system with respect to pure solvent, and the same holds for the thermodynamic quantities that will be computed in the following. We shall omit the "excess" specification for brevity.

In order to determine equilibrium properties, we have to minimize the grand potential with respect to pair probabilities. Notice that in principle we have taken into account two different site (marginal) PDs, even if the constraint Eq. (20) must hold. This is a trick, which allows us to write equations that preserve the symmetry of the pair PDs Eq. (21), and to avoid imposing the constraint explicitly. Nevertheless, we have to take into account other constraints. First of all, the pair PD must be (by definition) normalized:

$$\sum_{k=-2}^{2}\sum_{k'=-2}^{2} w_k w_{k'} p_{k,k'}^{(1,2)}=1. \tag{22}$$

Second, it must be taken into account that the states ±1 and ±2 have been distinguished to account for different pair energy contributions and the connectivity constraint, but, as previously mentioned, all local chain configurations must have the same probability, whence

$$p_s^{(i)}=p_{2s}^{(i)}, \tag{23}$$

for $s=\pm 1$ and $i=1,2$. Finally, we should take into account the constraint Eq. (8), to fix the PD of the quenched disorder.

We do so in an approximate way, by fixing the average concentration of the two monomer species on each site, that is, by imposing the constraint

$$\sum_{s=\pm 1} s p_s^{(i)} = m \sum_{s=\pm 1} p_s^{(i)}, \tag{24}$$

again for $s=\pm 1$ and $i=1,2$, where m is the quenched average defined by

$$m=\sum_{s=\pm 1} s P_s, \tag{25}$$

representing the relative concentration of H monomers with respect to P monomers. We do not need to impose the statistical independence of chemical species along the chain, because it is possible to verify that it comes out naturally as a consequence of the fact that any pair of chemically bonded monomers (H-H,P-P,H-P) yields the same energy (see Table II). Moreover, as previously mentioned, we do not even need to impose the symmetry constraints Eqs. (20) and (21), because they turn out to be automatically satisfied if the other constraints Eqs. (23) and (24) are imposed for both sites (1 and 2), thus giving rise to equations that are symmetric under exchange of k and k'.

It is possible to perform numerically a minimization of the grand potential ω [Eq. (18)], with the constraints Eqs. (22), (23), and (24), by means of an iterative method first proposed by Kikuchi [24]. Let us only notice that the method allows us to compute the equilibrium (pair and site) PDs $p_{k,k'}^{(1,2)}$ and $p_k^{(1)}=p_k^{(2)}$, which give information on the microscopic state of the system, as well as the equilibrium grand potential ω itself, from which all other thermodynamic quantities can be derived. We leave technical details of the method to the Appendix.

IV. NUMERICAL RESULTS

We now present the results of a calculation about an example of the class of systems introduced in Sec. II. We concentrate on the case in which $m=0$ (equal proportions of H and P monomers) and in which H monomers interact with one another with an energy $\epsilon_{+,+}\equiv-\epsilon$, with $\epsilon>0$. This simple model has been considered in the literature [14–16] to describe the effect of hydrophobicity on the behavior of H-P copolymers.

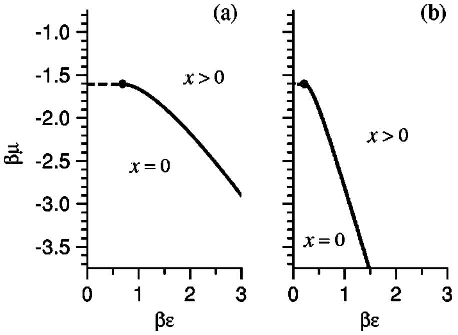


FIG. 1. Phase diagram in the plane $\beta\epsilon$-$\beta\mu$ for the three-dimensional cubic lattice $(q=6)$, in the cases $m=0$, equal proportions of H and P monomers (a), and $m=+1$, H monomers only (b). Dashed (solid) lines denote continuous (discontinuous) transitions. Circles denote tricritical points.

We compute, as an order parameter, the probability $x=1-p_0^{(i)}\equiv 1-p_0$ that a site is occupied by a monomer of any species, that is, the monomer volume fraction. Notice, by the way that, due to the homogeneity condition Eq. (20), from now on we shall always omit the PD superscripts. Given that the stable phase is the one with the minimum grand potential, it is possible to investigate the phase diagram of the system. This is displayed in Fig. 1(a) in the plane $\beta\epsilon$-$\beta\mu$ for the three-dimensional cubic lattice, corresponding to a coordination number $q=6$. We obtain a phase transition between a polymerized phase, in which $x>0$, and a nonpolymerized (pure solvent) phase, in which $x=0$. The transition is continuous for low values of $\beta\epsilon$ and discontinuous for higher values, and the two regions are separated by a tricritical point. The polymerized phase, in the limit of $\beta\mu$ tending to the transition line, represents an isolated chain in solution [11]: the continuous transition region $(x\rightarrow 0^+)$ corresponds to an extended (coil) state, while the coexistence region corresponds to a compact (globule) state. The tricritical point (Θ point) separates the two regimes. It is important to notice that this sort of picture implies that surface effects are neglected, that is, (when the polymer is in the globule state) the interaction of surface monomers with the solvent is not taken into account. This question will be addressed again in Sec. VI, where we discuss our results in comparison with models attempting to describe the effect of hydrophobicity in protein folding. In Fig. 1(b) the phase diagram is displayed for the homopolymer case (in which every monomer-monomer contact yields an energy $-\epsilon$). The phase diagram is qualitatively similar but the Θ transition is obtained for a lower value of $\beta\epsilon$, due to the fact that noninteracting monomers in the previous case favor the coil state.

We now analyze the most significant thermodynamic functions, namely, the Helmholtz free energy, internal energy, and entropy. Due to the fact that we always consider excess quantities with respect to the pure solvent, it turns out that ω, which we defined in the previous section, vanishes in the pure water phase, and on the transition line as well. As a consequence, it is possible to show by standard thermodynamics that μ at the transition line coincides with the Helmholtz free energy per monomer f in the infinite dilution limit. Remembering Eq. (18) it is easy to see that the internal energy per monomer u can be evaluated as

$$u=\tilde{\mu}+\frac{q}{2x}\sum_{k=-2}^{2}\sum_{k'=-2}^{2} w_k w_{k'} p_{k,k'} h_{k,k'} . \tag{26}$$

The entropy per monomer comes from $s/k_B=\beta(u-f)$. In Fig. 2 we report the Helmholtz free energy, internal energy, and entropy differences with respect to the expanded coil, as functions of $\beta\epsilon$. It can be shown analytically (we refer to Sec. V for the derivation) that in the coil phase the internal energy per monomer vanishes, whereas the entropy per monomer is

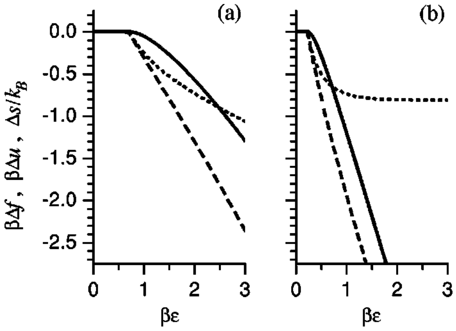


FIG. 2. Differences of Helmholtz free energy ($\beta\Delta f$, solid line), internal energy ($\beta\Delta u$, dashed line), and entropy ($\Delta s/k_B$, dotted line) per monomer of the globule phase with respect to the coil phase, as functions of $\beta\epsilon$. In (a) we have $m=0$, equal proportions of H and P monomers, and in (b) $m=+1$, H monomers only.

$$s=k_B\ln(q-1). \tag{27}$$

The former result means that a vanishing number of monomer-monomer contacts is realized, while the latter reflects the fact that the polymer chooses among $q-1$ equally probable directions at each step. Of course, this is a simplified representation of an expanded coil in the Bethe approximation. The internal energy and entropy in the coil phase can also be determined numerically in principle, by extrapolation as $\beta\mu$ tends to the second order transition line. Nevertheless, in this region the algorithm becomes slower and slower and numerical errors increase; hence the analytical results which we derive in Sec. V are helpful. In Fig. 2(a) it is possible to see that, as expected, internal energy and entropy decrease, as $\beta\epsilon$ increases, signaling that the system undergoes a collapse, increasing the number of contacts between "active" (H) monomers. It is interesting to compare the results with the purely H homopolymer case, reported in Fig. 2(b). It can be seen that both internal energy and entropy have a slower decrease in the heteropolymer case but, as far as entropy is concerned, this is true only near the transition. This suggests that the heteropolymer globule tends to be more ordered (to have a lower entropy) than the homopolymer one, due to "selective" monomer-monomer interactions.

In Fig. 3 we report some parameters, related to the equilibrium PD, that give some significant information about the local structure of the system. First of all we plot the site occupancy probability, equal to the monomer volume fraction x, vanishing in the coil phase (in the infinite dilution limit) and continuously increasing after the Θ transition. For the heteropolymer case the increase is slower, confirming the picture of a less abrupt transition. Secondly we present the monomer-monomer contact probability y, defined as the conditioned probability that a monomer (of any species) has a certain NN site occupied by a monomer (of any species) which is not consecutive along the chain:

$$y\equiv\frac{\sum_{s=\pm1}\sum_{s'=\pm1}p_{2s,2s'}}{\sum_{s=\pm1}p_{2s}}. \tag{28}$$

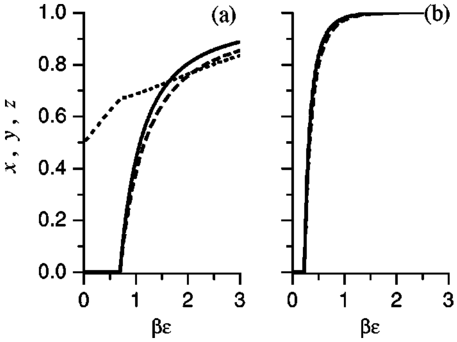


FIG. 3. Equilibrium average local structure properties: monomer volume fraction x (solid line), monomer-monomer contact probability y, as defined by Eq. (28) (dashed line), and, only in (a), H-H contact probability z, as defined by Eq. (29) (dotted line). In (a) we have $m=0$, equal proportions of H and P monomers, and in (b) $m=+1$, H monomers only.

The numerical value of y is not very different from the average density x, but the difference is slightly more evident in the heteropolymer case, and indeed it becomes more important as $\beta\epsilon$ increases, unlike in the homopolymer case. Due to the fact that this difference is an indicator of the correction of the Bethe approximation to the mean-field theory, a suggestion arises that the importance of the correction is greater for the heteropolymer case. Both the above quantities are indicators of the compactness of the polymer. For the heteropolymer, however, we need also an indicator of the order induced by the preference of H monomers to make contacts with one another. We choose the conditioned probability z that a monomer in contact with a H monomer is of the same species. In terms of pair PDs we have

$$z\equiv\frac{p_{+2,+2}}{\sum_{s'=\pm1}p_{+2,2s'}}. \tag{29}$$

In the coil phase region this quantity can be evaluated analytically, and turns out to be

$$z=\frac{1}{1+e^{-\beta\epsilon}P_-/P_+}. \tag{30}$$

We again refer to Sec. V for the derivation. It is worth noting that z is an increasing function of $\beta\epsilon$, and actually already starts increasing from the trivial value 1/2 (no monomer preference) in the coil phase. Moreover, the slope of the curve is lower in the globule phase, even if z is still increasing. Indeed, this is not surprising, and the physical meaning is that when the polymer is in a collapsed state it is more difficult for it to rearrange monomers of different species in order to satisfy energetically advantageous bonds. In a few words the model predicts that some rearrangements must take place already when the polymer is in an extended phase. This feature is easily caught by the pair approximation, but would be missed by a mean-field theory.

V. LANDAU EXPANSION

In this section we perform a Landau expansion of the grand potential ω [Eq. (18)] in powers of the monomer volume fraction x, which is the order parameter of the system. This is useful to determine analytically the second order transition and the tricritical point, which is quite a time-consuming task, by the numerical algorithm employed to perform the minimization of ω. First of all we have to write the grand potential as a function of only independent probability variables, which turn out to be $p \equiv p_{+1} + p_{-1} = p_{+2} + p_{-2}$, related to the monomer volume fraction by $x = (w_1 + w_2)p$, and $p_{2s,2s'}$, for $s,s' = \pm 1$. Notice that, according to Eq. (21), we have $p_{+2,-2} = p_{-2,+2}$, and hence we should consider only four independent variables. Nevertheless, we keep the above two variables distinct to simplify the notation. Making use of all the constraints Eqs. (21) and (22)–(24), of Eqs. (19), and of the fact that some pair states are not allowed because they violate connectivity constraints, one easily obtains all site and pair probabilities as functions of the independent variables:

$$p_0 = 1 - (w_1 + w_2)p, \tag{31}$$

$$p_s = p_{2s} = P_s p, \tag{32}$$

$$p_{0,0} = 1 - (w_1 + 2w_2)p + w_2^2 \sum_{s=\pm 1} \sum_{s'=\pm 1} p_{2s,2s'}, \tag{33}$$

$$p_{0,s} = p_{s,0} = 0, \tag{34}$$

$$p_{0,2s} = p_{2s,0} = P_s p - w_2 \sum_{s'=\pm 1} p_{2s,2s'}, \tag{35}$$

$$p_{s,s'} = p_{s',s} = P_s P_{s'} p / w_1, \tag{36}$$

$$p_{s,2s'} = p_{2s',s} = 0. \tag{37}$$

We can use Eq. (35), together with Eq. (26), to show that the present model can be mapped onto a model with monomer-monomer interactions only. The mapping results in the following substitutions, to be performed in Table II,

$$\epsilon_{s,s'} \to \epsilon_{s,s'} - \gamma_s - \gamma_{s'},$$

$$\gamma_s \to 0,$$

$$\mu \to \mu - (q-2)\langle \gamma_s \rangle_s, \tag{38}$$

where $s,s' = \pm 1$ and

$$\langle \gamma_s \rangle_s \equiv \sum_{s=\pm 1} P_s \gamma_s. \tag{39}$$

In the following we shall actually study such a simplified model. We then substitute Eqs. (31)–(37) in the expression of the grand potential Eq. (18), taking the derivatives with respect to the independent variables:

$$\frac{\partial \beta\omega}{\partial x} = -[\beta\mu + \ln(q-1)] + (q-1)\ln\frac{p_0}{p_{0,0}} - (q-2)\sum_{s=\pm 1} P_s \ln\frac{P_s p}{p_{0,2s}}, \tag{40}$$

$$\frac{\partial \beta\omega}{\partial p_{2s,2s'}} = \frac{q}{2} w_2^2 \left(\beta\epsilon_{s,s'} + \ln\frac{p_{0,0} p_{2s,2s'}}{p_{0,2s} p_{0,2s'}} \right). \tag{41}$$

Let us notice that, substituting Eqs. (31)–(37) into Eq. (18) except in the arguments of the logarithms, and taking into account Eqs. (40), (41), it is possible to show that

$$\beta\omega = \frac{q}{2}\ln p_{0,0} - (q-1)\ln p_0 + x\frac{\partial \beta\omega}{\partial x} + \sum_{s=\pm 1}\sum_{s'=\pm 1} p_{2s,2s'} \frac{\partial \beta\omega}{\partial p_{2s,2s'}}, \tag{42}$$

which will be useful in the following. We now set to zero the derivatives Eqs. (41), obtaining some of the necessary conditions for the minimum of the grand potential,

$$p_{2s,2s'} = e^{-\beta\epsilon_{s,s'}} \frac{p_{0,2s} p_{0,2s'}}{p_{0,0}}. \tag{43}$$

This set of equations defines, through Eqs. (33) and (35), four implicit functions $p_{2s,2s'}(p)$, that allow one in principle to determine the grand potential as a function of p only. The minimum condition is then obtained by setting to zero the right hand side of Eq. (40). The implicit functions $p_{2s,2s'}(p)$ are not easy to determine, due to the fact that Eqs. (43) are coupled with one another. Anyway we are interested only in the asymptotic behavior for $p \to 0$, which is easily verified to be

$$p_{2s,2s'} = e^{-\beta\epsilon_{s,s'}} P_s P_{s'} p^2 + o(p^2). \tag{44}$$

Let us notice that Eq. (44) is sufficient to prove Eq. (30), taking the limit $p \to 0$. Moreover, substituting into Eqs. (33) and (35), we have

$$p_{0,0} = 1 - (w_1 + 2w_2)p + \langle e^{-\beta\epsilon_{s,s'}} \rangle_{s,s'} w_2^2 p^2 + o(p^2), \tag{45}$$

$$p_{0,2s} = P_s p [1 - \langle e^{-\beta\epsilon_{s,s'}} \rangle_{s'} w_2 p + o(p)], \tag{46}$$

where

$$\langle e^{-\beta\epsilon_{s,s'}} \rangle_{s'} \equiv \sum_{s'=\pm 1} P_{s'} e^{-\beta\epsilon_{s,s'}}, \tag{47}$$

$$\langle e^{-\beta\epsilon_{s,s'}} \rangle_{s,s'} \equiv \sum_{s=\pm 1}\sum_{s'=\pm 1} P_s P_{s'} e^{-\beta\epsilon_{s,s'}}. \tag{48}$$

Finally, substituting Eqs. (45), (46), and (31) into Eq. (40) and hence into Eq. (42), where the last term is obviously $o(p^2)$, we obtain

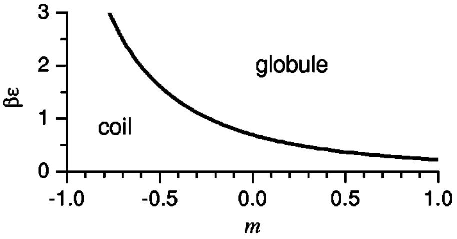


FIG. 4. Θ transition line: $(\beta\epsilon)_\Theta$ as a function of m.

$$\beta\omega = -[\beta\mu + \ln(q-1)]x + \frac{q-2}{2q} \times [(q-1) - (q-2)\langle e^{-\beta\epsilon_{s,s'}}\rangle_{s,s'}]x^2 + o(x^2), \tag{49}$$

which is just the required expansion.

It is easy to see that, when the monomer volume fraction x vanishes, the grand potential is zero, i.e., it is equal to that of pure solvent, as one could expect. Moreover, one can observe that, if the second order term is positive, a continuous transition can take place when the first order term changes sign. This leads us to locate the continuous transition at

$$\beta\mu = -\ln(q-1). \tag{50}$$

Due to the fact that μ at the transition coincides with the Helmholtz free energy per monomer of the isolated polymer in solution, we can also verify that it is made up of the only entropic term Eq. (27), because Eq. (44) implies that the internal energy per monomer is $O(x^2)/x = O(x)$ in the limit $x \to 0$ and hence vanishes in the coil phase. We can also observe that, if on the contrary the second order term is negative, only a discontinuous transition can take place, in a region where the first order term is positive, although we cannot locate it analytically. Anyway, we see that a vanishing second order coefficient

$$\langle e^{-\beta\epsilon_{s,s'}}\rangle_{s,s'} = \frac{q-1}{q-2} \tag{51}$$

determines the separation between the continuous and discontinuous transition regions, i.e., the Θ point. We can solve this equation for the case of our simple test model, analyzing the behavior of the Θ transition (namely, of the value of $\beta\epsilon$ where the transition takes place) as a function of the parameter m, related to the fraction of H monomers in the chain (see Fig. 4). For $m = +1$ only H monomers are present and one obtains

$$(\beta\epsilon)_\Theta = \ln\frac{q-1}{q-2} \tag{52}$$

($\approx$ 0.2231 for $q = 6$). Upon increasing the fraction of P monomers, $(\beta\epsilon)_\Theta$ increases as well, and hence the globule turns out to be less and less stable. The curve asymptotically tends to infinity for $m \to -1$, meaning that the purely "hydrophilic" polymer has no collapsed state, as could be expected.

VI. DISCUSSION AND CONCLUSIONS

In this paper we have generalized the pair approximation scheme of the CVM to a lattice random copolymer model, with particular attention to the case of H-P copolymer. We have also analyzed thermodynamic functions (free energy, internal energy, entropy) and parameters related to the equilibrium (pair) PD in a particularly simple case, previously investigated by Dill [15]. It is worth remarking on some interesting issues.

First of all we observed that the entropy of the collapsed state is considerably lower when we consider the case of H-P copolymer, with respect to the case of a purely H homopolymer. This is a significant effect, which can be ascribed to the "selective" interaction between monomers of a certain species, which makes the polymer collapse into a more ordered state. Such an entropy lowering was also observed in Ref. [19] in the case of a weakly H polymer (and actually we expect that this effect will be further enhanced by an increasing concentration of P monomers), but in that case it was accompanied by an unusual first order collapse, whereas in our case the transition is continuous.

In addition to the thermodynamic quantities, we have also analyzed in some detail pair correlations at thermodynamic equilibrium. The most interesting parameter is perhaps that we have denoted by z, which describes quantitatively the preference of H monomers to be in contact with monomers of the same species. It turns out that it is an increasing function of the interaction parameter $\beta\epsilon$ in the coil phase also. On the one hand this means that the coil phase is actually nontrivial, because some ordering is induced by the preference among H monomers, even if the polymer is globally in an extended state. On the other hand it suggests that the Bethe approximation is able to catch this effect, which would not be taken into account in the hypothesis of neglecting pair correlations.

An important part of the paper has been devoted to the Landau expansion, which gives information about the coil phase and the Θ point analytically. We compare the latter result with the original work by Aguilera-Granja and Kikuchi [6]. They considered a homopolymer model with a monomer-solvent interaction energy J, which in our notation means $\gamma_+ = J$ and $\gamma_- = \epsilon_{s,s'} = 0$. Using Eqs. (38) and (52), the Flory parameter at the Θ point turns out to be

$$\chi_\Theta = (\beta q J)_\Theta = \frac{q}{2}\ln\frac{q-1}{q-2}. \tag{53}$$

The numerical values obtained by this formula are compared in Table III with those obtained numerically in Ref. [6]. The small discrepancy for low values of q can probably be ascribed to the fact that in the cited paper numerical values are computed at finite chain length. Moreover, it is easy to see that

$$\lim_{q\to\infty} \chi_\Theta = \frac{1}{2}, \tag{54}$$

TABLE III. Comparison between the Θ points (in terms of Flory's parameter χ) obtained in the present work, Eq. (53), and in Ref. [6], for the H homopolymer case and for different values of the coordination number q.

q	χ_Θ: Eq. (53)	χ_Θ: Ref. [6]
4	0.8109	0.77
6	0.6694	0.675
8	0.6166	0.61
12	0.5719	0.57

confirming the fact that Flory's theory [10] gives the correct result in the infinite coordination limit [6].

We also compare our results with those obtained by Dill [15], who worked out a mean-field-like treatment of an analogous model on a "spherical" lattice. To do so we recall that in the cited paper the stability of a globular protein is estimated through the free energy difference between the "ordered" collapsed state, representing the folded state, and a "randomly collapsed" state, representing the unfolded one. We assume that in the framework of our model, which does not possess an underlying "globular" geometry, the randomly collapsed state could be reasonably represented by the ordinary collapsed state of an "equivalent homopolymer" in which the monomer-monomer interaction equals ϵP_+^2, that is, the H-H contact energy reduced by a factor which takes into account the random arrangement of monomers. In Fig. 5 we report the free energy difference with respect to such a "homopolymer" system, for different values of $\beta\epsilon$, as a function of the parameter m, related to the H monomer concentration. Of course it turns out that a stronger interaction corresponds to a higher stability, but the most important feature is that a concentration of maximum stability is observed, as predicted by Dill's model. Above and below that concentration the "folded" state becomes less stable. Let us notice that in our treatment, due to the the fact that the surface effect is neglected (as mentioned in Sec. IV), there is no possibility of describing a global rearrangement of H monomers in the core and of hydrophilic monomers outside. Only local rearrangements are taken into account and this is probably the reason why the maximum stability effect can be observed even in the infinite length limit, unlike in Dill's model. Other discrepancies can be observed at high concentrations of P monomers (low m values), where we never have an unfolded state more stable than the folded one, due to the fact that the "equivalent homopolymer" always reaches the Θ point at higher m values than the copolymer.

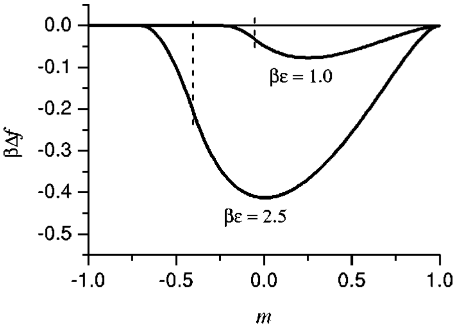


FIG. 5. Difference of Helmholtz free energy per monomer ($\beta\Delta f$) with respect to the "equivalent homopolymer" as a function of m, for different values of $\beta\epsilon$. Thin dashed lines mark the Θ transitions of the "equivalent homopolymer." The Θ transitions of the copolymer take place where the free energy differences vanish.

Finally, we would like to note that one might expect the possibility of taking into account at least local (pair) correlations along the sequence, by means of the Bethe approximation. This might be of some relevance, because in the framework of the H-P model important correlations have recently been observed [25] in sequences properly designed to have globular conformations with H monomers shielded by P ones. Unfortunately, it turns out that imposing pair correlations between contiguous monomers is simply equivalent (in our scheme) to a chemical potential shift, with no effect on the single polymer properties. It might be possible to take such correlations into account by considering larger basic clusters in the spirit of the CVM. This goes beyond the scope of the present paper and will be the subject of a future work.

ACKNOWLEDGMENTS

I am indebted to P. Bruscolini, C. Buzano, and A. Pelizzola for many helpful suggestions and discussions. This work was partially supported by the Italian MURST, through the research project "Heteropolymers and Proteins."

APPENDIX: NUMERICAL MINIMIZATION OF ω

As mentioned in Sec. III, in this appendix we give the technical details of the numerical method employed to find constrained minima of the grand potential with respect to pair probabilities $p_{k,k'}^{(1,2)}$. According to the Lagrange multiplier method, such minima can be found as the free minima of another functional, namely,

$$\beta\tilde{\omega}=\beta\omega-\frac{q}{2}\left(\sum_{k=-2}^{2}\sum_{k'=-2}^{2} w_k w_{k'} p_{k,k'}^{(1,2)}-1\right)\mathcal{L} -\frac{q}{2}\sum_{i=1,2}\sum_{s=\pm 1}\left[(p_s^{(i)}-p_{2s}^{(i)})\mathcal{K}_s^{(i)}+(s-m)p_s^{(i)}\mathcal{J}^{(i)}\right], \tag{A1}$$

with respect both to the pair probabilities and the Lagrange multipliers $\mathcal{J}^{(i)}$, $\mathcal{K}_s^{(i)}$, and $\mathcal{L}$. Setting to zero the partial derivatives of $\beta\tilde{\omega}$ with respect to $p_{k,k'}^{(1,2)}$ gives rise to a system of equations, easily written in fixed point form

$$p_{k,k'}^{(1,2)}=\mathcal{W}\exp(-\beta h_{k,k'}+\mathcal{V}_k^{(1)}+\mathcal{V}_{k'}^{(2)}+\mathcal{U}_k^{(1)}+\mathcal{U}_{k'}^{(2)}) \times(p_k^{(1)}p_{k'}^{(2)})^{(q-1)/q}, \tag{A2}$$

where the unknown parameters $\mathcal{U}_k^{(i)}$, $\mathcal{V}_k^{(i)}$, and $\mathcal{W}$ are related to the Lagrange multipliers in the following way:

$$\mathcal{U}_k^{(i)}=\mathcal{J}^{(i)}\delta_{|k|,1}(k-m), \tag{A3}$$

$$\mathcal{V}_k^{(i)} = \mathcal{K}^{(i)}_{\mathrm{sgn}\,k} \frac{\delta_{|k|,1} - \delta_{|k|,2}}{w_k}, \tag{A4}$$

$$\mathcal{W} = \exp\left(\mathcal{L} + \frac{q-2}{q} \right) \tag{A5}$$

(sgn denotes the *sign function*, returning $+1$, -1, or 0 if its argument is, respectively, greater than, less than, or equal to zero). The numerical problem is solved as follows. Equation (A2) is used as elementary iteration of a fixed point method, usually called the natural iteration method [23]. Nonetheless, we need to determine $\mathcal{U}_k^{(i)}$, $\mathcal{V}_k^{(i)}$, and $\mathcal{W}$ at each iteration, in order to satisfy the constraints. We do so by a nested iterative method, involving the following steps.

(1) Set $\mathcal{K}_s^{(i)} = 0$, $\mathcal{J}^{(i)} = 0$, and $\mathcal{W} = 1$ ($\mathcal{W}$ is a simple normalization constant).

(2) Compute a trial "next step" estimate of $p_{k,k'}^{(1,2)}$ by means of Eq. (A2).

(3) Adapt the Lagrange multipliers by the iterative equations

$$\hat{\mathcal{K}}_s^{(i)} = \mathcal{K}_s^{(i)} - \alpha \frac{w_1 w_2}{w_1 + w_2} \ln \frac{p_s^{(i)}}{p_{2s}^{(i)}}, \tag{A6}$$

$$\hat{\mathcal{J}}^{(i)} = \mathcal{J}^{(i)} - \alpha \left(\frac{1}{2} \log \frac{p_{+1}^{(i)}}{p_{-1}^{(i)}} - \tanh^{-1} m \right), \tag{A7}$$

where carets denote the new estimates and α is a "relaxation" parameter, chosen empirically ($\alpha \approx 0.4$ is usually a good choice).

(4) If the distances $|\hat{\mathcal{K}}_s^{(i)} - \mathcal{K}_s^{(i)}|$ and $|\hat{\mathcal{J}}^{(i)} - \mathcal{J}^{(i)}|$ are larger than some tolerance (i.e., if the estimate of $p_{k,k'}^{(1,2)}$ does not satisfy the constraints within the tolerance), go to step 2.

(5) Determine $\mathcal{W}$ by normalization.

The nested iteration method, first applied by Kikuchi [24], proves to converge quite easily for this problem. Let us notice that Eqs. (A6) and (A7) are derived in a nonrigorous way, although following precise criteria. With $i = 1{,}2$ and $s = \pm 1$, let $\hat{\mathcal{K}}_s^{(i)}$ and $\mathcal{K}_s^{(i)}$ be, respectively, the "new" and the "old" estimates of the Lagrange multipliers used to impose the constraint Eq. (23), and $\hat{p}_s^{(i)}$, and $p_s^{(i)}$ ($\hat{p}_{2s}^{(i)}$ and $p_{2s}^{(i)}$) the corresponding estimates of the site probabilities, evaluated by Eq. (A2) and Eqs. (19). Taking into account also Eq. (A4), it is easy to show that

$$\hat{p}_s^{(i)} = \frac{\exp(\hat{\mathcal{K}}_s^{(i)}/w_1)}{\exp(\mathcal{K}_s^{(i)}/w_1)} p_s^{(i)}, \tag{A8}$$

$$\hat{p}_{2s}^{(i)} = \frac{\exp(-\hat{\mathcal{K}}_s^{(i)}/w_2)}{\exp(-\mathcal{K}_s^{(i)}/w_2)} p_{2s}^{(i)}. \tag{A9}$$

Of course we would like the new estimates of site probabilities to satisfy the constraint Eq. (23) and hence we impose

$$\hat{p}_s^{(i)} = \hat{p}_{2s}^{(i)}. \tag{A10}$$

We then easily obtain Eq. (A6) with $\alpha = 1$. In this way the iterative procedure usually does not converge, and this is the reason why we introduce the relaxation parameter $\alpha < 1$, adjusted empirically. A similar derivation can be performed for Eq. (A7). By means of Eqs. (A2), (19), and (A3) we can write, with analogous meaning of the symbols,

$$\hat{p}_s^{(i)} = \frac{\exp[\hat{\mathcal{J}}^{(i)}(s-m)]}{\exp[\mathcal{J}^{(i)}(s-m)]} p_s^{(i)}. \tag{A11}$$

Hence, imposing the constraint Eq. (24), that is,

$$\sum_{s=\pm 1} s \hat{p}_s^{(i)} = m \sum_{s=\pm 1} \hat{p}_s^{(i)}, \tag{A12}$$

one easily obtains Eq. (A7) with $\alpha = 1$.

It is finally possible to show that, once the iterative procedure has reached convergence, one has

$$\beta \omega = \frac{q}{2} \ln \mathcal{W}, \tag{A13}$$

which allows us to evaluate easily the equilibrium grand potential per site. Let us also notice that, if we start the procedure with a translationally invariant (that is, independent of i) PD, and with guess values of the Lagrange multipliers independent of i as well, the whole iterative procedure is independent of the site index. This means that actually it is not necessary to define two different sets of Lagrange multipliers (for $i = 1{,}2$), but only one is sufficient. We can then remove the superscripts from all Lagrange multipliers and related parameters.

[1] H. A. Bethe, Proc. R. Soc. London, Ser. A **150**, 552 (1935).

[2] E. A. Guggenheim, *Mixtures* (Oxford University Press, London, 1952).

[3] R. Kikuchi, Phys. Rev. **81**, 988 (1951).

[4] D. M. Burley, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 2, Chap. 9.

[5] G. An, J. Stat. Phys. **52**, 727 (1988).

[6] F. Aguilera-Granja and R. Kikuchi, Physica A **176**, 514 (1991).

[7] F. Aguilera-Granja and R. Kikuchi, Physica A **182**, 331 (1992).

[8] F. Aguilera-Granja and R. Kikuchi, Physica A **189**, 81 (1992); **189**, 108 (1992); **195**, 53 (1993).

[9] P. J. Flory, J. Chem. Phys. **10**, 51 (1942).

[10] P. J. Flory, J. Chem. Phys. **12**, 425 (1944).

[11] S. Lise, A. Maritan, and A. Pelizzola, Phys. Rev. E **58**, R5241 (1998).

[12] J. des Cloizeaux and G. Jannink, *Polymers in Solution: Their Modelling and Structure* (Clarendon Press, Oxford, 1990).

[13] C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, England, 1998).
[14] K. A. Dill, Biochemistry **29**, 7133 (1990).
[15] K. A. Dill, Biochemistry **24**, 1501 (1985).
[16] K. A. Dill, D. O. V. Alonso, and K. Hutchinson, Biochemistry **28**, 5439 (1989).
[17] V. Shahrezaei, N. Hamedani, and M. R. Ejtehadi, Phys. Rev. E **60**, 4629 (1999).
[18] V. Shahrezaei and M. R. Ejtehadi, J. Chem. Phys. **113**, 6437 (2000).
[19] T. Garel, L. Leibler, and H. Orland, J. Phys. II **4**, 2139 (1994).
[20] A. Moskalenko, Yu. A. Kutznetsov, and K. A. Dawson, J. Phys. II **7**, 409 (1997).
[21] T. Morita, Prog. Theor. Phys. Suppl. **80**, 103 (1984).
[22] C. Buzano, A. Maritan, and A. Pelizzola, J. Phys.: Condens. Matter **6**, 327 (1994).
[23] R. Kikuchi, J. Chem. Phys. **60**, 1071 (1974).
[24] R. Kikuchi, J. Chem. Phys. **65**, 4545 (1976).
[25] E. N. Govorun *et al.*, e-print cond-mat/0102180 v2.