

## Gaussian self-consistent method for the kinetics of heteropolymers: A direction in studying the protein folding problem

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

*Theory Group, Centre for Soft Condensed Matter and Biomaterials, Department of Chemistry, University College Dublin,  
Dublin 4, Ireland*

(Received 16 October 1995)

We develop a Gaussian self-consistent method for the study of equilibrium and kinetics of conformational transitions of arbitrary heteropolymers in dilute solution. It is discovered that certain chain sequences possess an additional symmetry that leads to reduction of the number of dynamical variables and simplification of the equations. As an application of our general method we consider the problem for a periodic ring heteropolymer with the sequence  $ab$  constituted by monomers of two types with different second virial coefficients. We present numerical results for equilibrium and kinetics of this system on that subspace of the phase diagram where one group of monomers can become hydrophobic. We find an interesting physical phenomenon of a phase separation that accompanies the coil-to-globule transition. We believe that our approach may shed light on the fundamental problem of protein folding; some of the results seem encouraging.

PACS number(s): 36.20.-r, 87.15.-v

### I. INTRODUCTION

For many years considerable efforts have been directed at achieving an understanding of the collapse transition of homopolymers in dilute solution. Numerous studies have been carried out on the equilibrium phenomena [1–5] and lately there has been interest in the kinetics of this process [6–9]. The equilibrium theory is quite well understood and we have recently argued that homopolymer kinetics is also now a relatively complete story [10].

Apart from the intrinsic interest in the homopolymer problem there has long been an underlying belief that such studies would lead us closer to an understanding of the phenomena of biopolymer folding or compaction as observed in proteins, DNA, and other molecules [11]. While this is partly true, it is our belief that the main contribution of these works [12] is in providing a basis from which to study the collapse kinetics of heteropolymers. Thus it is widely accepted that among the essential features required in a biopolymer model is the presence of “frustration” due to a combination of monomer units with some type of opposing tendency to associate or repel each other, along with some connectivity constraint.

A logical consequence of this reasoning is that we should now be in a position to apply the machinery that proved successful in the kinetic studies of homopolymer to, for example, periodic heteropolymers. A simple example would be a copolymer composed of hydrophilic and hydrophobic units with some periodic chemical sequence along the chain. However, it transpires that simulation methods for such frustrated problems are very difficult, and it is unclear if there is a real possibility of deducing the physical laws on this basis [13]. However, one of the approaches we have developed, a sort of time-dependent mean-field method, seems to be more promising.

Indeed, we shall here argue that the extension of that

method to periodic heteropolymers allows us to capture the essential features of kinetics of polymers that possess frustration. The presence of frustration in the system is expected to lead to new kinetic phenomena absent in the homopolymer.

In that more simple case we have identified three distinct regimes. The first describes formation of a periodic necklace of locally collapsed clusters along the chain growing at the expense of their chain neighbors. This process, although it originates in a local cluster growth mechanism, is in essence a collective phenomenon akin to the spinodal decomposition but here in the internal metric of the chain. The second regime describes coarsening of the clusters and may be viewed as a constrained form of the Lifshitz-Slyozov mechanism. Importantly, it is dominated by the behavior of a nearly ideal chain of locally collapsed globules. The third stage is a slow relaxational process of shape optimization and compaction of a single globule. We have given concrete analytical laws governing all three stages [9,10].

However, we know that, at equilibrium, the copolymer typically has a tendency to expel its hydrophilic units to the exterior, with hydrophobic units moving to the interior. This may be viewed as some sort of constrained phase separation and it is of considerable interest to know if the overall scheme of homopolymer collapse outlined above is retained, with an additional stage added, or whether some more complex behavior emerges. It is also important to deduce the concrete laws of kinetics for this situation, for we have every reason to believe that these may then represent a much better description of biopolymer folding than the homopolymer case.

The remainder of this paper is organized as follows. Firstly, we outline the general method to deal with arbitrary heteropolymers. We then find that for a broad class of sequences there is a symmetry in the metric of the chain. This symmetry yields certain identities that

allow us to simplify the basic equations. As an example we apply the method to a simple binary ( $ab$ ) periodic sequence. The self-consistency equations are studied numerically and results presented.

We emphasize that the method we described is completely general; any arbitrary heteropolymer, with either ring or open topology, can be studied. In addition, though we here use Gaussian-type mean-field theory, the symmetry considerations are quite general and could be implemented in a more detailed theory of an interacting chain.

Finally, we point out that though the problem we study has direct relevance to biological issues, it has fundamental significance from the point of view of statistical mechanics. Thus, both fractal dimension and a density order parameter are involved in the transition, a novel feature in the field of statistical mechanics.

## II. THE METHOD

Technically it is more convenient to consider ring heteropolymers possessing a periodic structure. More precisely we assume that the chain is constituted from monomers of several different types in such a manner that they form a certain fundamental sequence (block) that is periodically repeated  $M$  times along the chain. Let us denote by  $\mathbf{x}_n^A$  the spatial position of the  $A$ th element in the  $n$ th block. Upper case indices take values  $A = 0, \dots, K-1$ , where  $K$  is the length of a block, and lower case indices take values  $n = 0, \dots, M-1$ , where  $M$  is the number of blocks in the chain. Evidently, the total number of beads is equal to  $N = KM$ .

Now, consider the Langevin equation, which, in our notation, is written

$$\zeta_b \dot{\mathbf{x}}_n^A = -\kappa \sum_{n'A'} J_{nn'}^{AA'} \mathbf{x}_{n'}^{A'} - \frac{\partial V}{\partial \mathbf{x}_n^A} + \boldsymbol{\eta}_n^A, \quad (1)$$

$$V = \sum_{L=2}^{\infty} \sum_{B_1 \dots B_L} u_{B_1 \dots B_L}^{(L)} \sum_{m_1 \dots m_L} \prod_{i=1}^{L-1} \delta(\mathbf{x}_{m_i}^{B_i} - \mathbf{x}_{m_{i+1}}^{B_{i+1}}), \quad (2)$$

$$\langle \boldsymbol{\eta}_n^A(t) \boldsymbol{\eta}_{n'}^{A'}(t') \rangle = 2k_B T \zeta \delta_{nn'} \delta^{AA'} \delta(t-t'), \quad (3)$$

with  $(m_1, B_1) \neq (m_{i+1}, B_{i+1})$ . The matrix  $J_{nn'}^{AA'}$  of spring constants may be described as follows. Its nonzero elements can be written in the form

$$J_{n,n}^{A,A} = 2, \quad J_{n,n}^{A,A+1} = J_{n,n}^{A,A-1} = -1, \quad (4)$$

$$n = 0, \dots, M-1, \quad A = 0, \dots, K-1,$$

where we have assumed the following convention:

$$J_{n,n}^{K-1,K} \equiv J_{n,n+1}^{K-1,0} = -1, \quad (5)$$

$$J_{n,n}^{0,-1} \equiv J_{n,n-1}^{0,K-1} = -1. \quad (6)$$

There are additional cyclicity properties in the chain index,

$$J_{M-1,M}^{K-1,0} = J_{M-1,0}^{K-1,0} = -1, \quad (7)$$

$$J_{0,-1}^{0,K-1} = J_{0,M-1}^{0,K-1} = -1. \quad (8)$$

One important observation is that, if one is restricted to the  $A$ th elements of blocks, such a subchain still possesses the translational invariance along its members. Thus the translation operator may be written as a multiplication in the space of the Fourier modes

$$\mathbf{x}_q^A = \frac{1}{M} \sum_{n=0}^{M-1} f_n^{(q)} \mathbf{x}_n^A, \quad f_n^{(q)} = \exp(i2\pi nq/M). \quad (9)$$

After the Fourier transformation the Langevin equation (1) and (3) becomes

$$\zeta \dot{\mathbf{x}}_q^A = -k \sum_{A'} J_q^{AA'} \mathbf{x}_q^{A'} - \frac{\partial V}{\partial \mathbf{x}_q^A} + \boldsymbol{\eta}_q^A, \quad (10)$$

$$\langle \boldsymbol{\eta}_q^A(t) \boldsymbol{\eta}_{q'}^{A'}(t') \rangle = 2k_B T \zeta \delta_{-qq'} \delta^{AA'} \delta(t-t'), \quad (11)$$

where  $\zeta = M\zeta_b$ ,  $k = M\kappa$ , and the spring matrix is diagonal in  $q$ ,

$$J_q^{AB} = 2\delta^{A,B} - \delta^{A,B+1} - \delta^{A+1,B} - f_q \delta^{A,0} \delta^{B,K-1} - f_q^* \delta^{A,K-1} \delta^{B,0}, \quad (12)$$

with  $f_q \equiv f_1^{(q)}$ . It is simple to extend the Gaussian self-consistent method [9,14,15] by choosing the Gaussian stochastic ensemble,

$$\zeta \dot{\mathbf{x}}_q^A(t) = - \sum_{A'} \Delta V_q^{AA'}(t) \mathbf{x}_q^{A'} + \boldsymbol{\eta}_q^A(t), \quad (13)$$

with a nondiagonal effective potential. Formal integration of this equation yields

$$\mathbf{x}_q(t) = G_q^V(t,0) \bullet \mathbf{x}_q(0) + \frac{1}{\zeta} \int_0^t dt' G_q^V(t,t') \bullet \boldsymbol{\eta}_q(t'). \quad (14)$$

Here the bullet means matrix multiplication and the Green function is the matrix time-ordered exponential,

$$G_q^V(t,t') = T \exp \left( -\frac{1}{\zeta} \int_{t'}^t dt'' \Delta V_q(t'') \right). \quad (15)$$

Using formula (14) one can calculate the average

$$\frac{1}{3} \langle \mathbf{x}_{-q}^{A'}(t) \boldsymbol{\eta}_q^A(t) \rangle = 2k_B T \delta^{AA'}. \quad (16)$$

Applying the technique described in Refs. [9,15] and Eqs. (10)–(13) one finds that the equal-time correlation function,

$$\mathcal{F}_q^{A'A}(t) = \frac{1}{3} \langle \mathbf{x}_{-q}^{A'}(t) \mathbf{x}_q^A(t) \rangle, \quad (17)$$

satisfies a nonlinear differential equation,

$$\zeta \dot{\mathcal{F}}_q^{AA'}(t) = 2k_B T \delta^{AA'} - k \sum_{A''} (J_q^{AA''} \mathcal{F}_q^{A'A''} + J_{-q}^{A'A''} \mathcal{F}_{-q}^{AA''}) + \sum_{L=2}^{\infty} \sum_{B_1 \dots B_L} \hat{u}_{B_1 \dots B_L}^{(L)} \sum_{m_1 \dots m_L} (\det \Delta^{(L-1)})^{-5/2} \sum_{i,j=1}^{L-1} \bar{\Delta}_{ij}^{(L-1)} \sum_{A''} (d_{ij}^{(q)AA''} \mathcal{F}_q^{A'A''} + d_{ij}^{(-q)A'A''} \mathcal{F}_{-q}^{AA''}), \quad (18)$$

where  $\hat{u}_{B_1 \dots B_L}^{(L)} = (2\pi)^{-3(L-1)/2} u_{B_1 \dots B_L}^{(L)}$  and  $\Delta^{(N-1)}$  is the matrix of size  $L-1$  with the elements

$$\Delta_{ij}^{(L-1)} = D_{m_1 m_{i+1}, m_1 m_{j+1}}^{B_1 B_{i+1}, B_1 B_{j+1}}, \quad (19)$$

the cofactors being denoted by  $\bar{\Delta}_{ij}^{(L-1)}$ . We have also used the contracted notation

$$d_{ij}^{(q)AA'} = d_{m_1 m_{i+1}, m_1 m_{j+1} B_1 B_{i+1}, B_1 B_{j+1}}^{(q)AA'}. \quad (20)$$

These four-body coefficients are, in turn, related to  $f_m^{(q)}$  by

$$d_{mm', nn', BB', CC'}^{(q)AA'} = c_{mm' BB', CC'}^{(-q)A} c_{nn' CC'}^{(q)A'}, \quad (21)$$

$$c_{mm' BB'}^{(q)A} = f_m^{(-q)} \delta_B^A - f_{m'}^{(-q)} \delta_{B'}^A,$$

and the quantities

$$D_{mm', nn'}^{BB', CC'} = \frac{1}{3} \langle (\mathbf{x}_m^B - \mathbf{x}_{m'}^{B'}) (\mathbf{x}_n^C - \mathbf{x}_{n'}^{C'}) \rangle \quad (22)$$

are the Fourier transforms of the internal modes  $\mathcal{F}_q^{AA'}$ ,

$$D_{mm', nn'}^{BB', CC'} = \sum_{q, AA'} d_{mm', nn', BB', CC'}^{(q)AA'} \mathcal{F}_q^{AA'}. \quad (23)$$

The four-body functions may be reduced to simpler two-body ones defined by

$$d_{mm' BB'}^{(q)} \equiv d_{mm', mm' BB', BB'}^{(q)}, \quad D_{mm'}^{BB'} \equiv D_{mm', mm'}^{BB', BB'}, \quad (24)$$

using the following relations:

$$d_{mm', nn', BB', CC'}^{(q)} = \frac{1}{2} (d_{mn' BC'}^{(q)} + d_{m' n B' C}^{(q)} - d_{mn BC}^{(q)} - d_{m' n' B' C'}^{(q)}). \quad (25)$$

Analogous relations hold for  $D_{mm' nn'}^{BB' CC'}$ . The foregoing treatment may be generalized to include the hydrodynamic interaction and the results are described in Appendix A.

It is interesting to note that, in comparison to a homopolymer, the modes  $\mathcal{F}_0^{AA'}$  describe conformational motions as well as diffusion of the center of mass. The radius of gyration may be expressed through the Fourier modes as follows:

$$R_g^2(t) = \frac{1}{K} \sum_{q \neq 0, A} \mathcal{F}_q^{AA}(t) + \frac{1}{2K^2} \sum_{AA'} Z^{AA'}, \quad (26)$$

where we have introduced  $K(K-1)/2$  linear combina-

tions of the zero modes [16],

$$Z^{AA'} = \mathcal{F}_0^{AA} + \mathcal{F}_0^{A'A'} - 2\mathcal{F}_0^{AA'} = \frac{1}{3} \langle (\mathbf{R}^A - \mathbf{R}^{A'})^2 \rangle, \quad (27)$$

$$\mathbf{R}^A \equiv \frac{1}{M} \sum_{n=0}^{M-1} \mathbf{x}_n^A. \quad (28)$$

These quantities are the mean squared distances between the center of mass of the  $A$ th and  $A'$ th monomers in all blocks. Note that these combinations,  $Z^{AA'}$ , of the zero modes contribute to the spatial correlations in (23). The diffusion mode describes the center of mass of the whole polymer and is given by the formula

$$W = \sum_{AA'} \mathcal{F}_0^{AA'} = \frac{1}{3} \left\langle \left( \sum_A \mathbf{R}^A \right)^2 \right\rangle. \quad (29)$$

The remaining  $K-1$  independent linear combinations,

$$Y^A = \mathcal{F}_0^{00} - \mathcal{F}_0^{AA} = \frac{1}{3} \langle (\mathbf{R}^0)^2 - (\mathbf{R}^A)^2 \rangle, \quad A = 1, \dots, K-1, \quad (30)$$

satisfy homogeneous differential equations, which thus leads to a trivial solution if the initial condition is equal to zero.

Now, let us write down formulas for some important observables. The average energy is

$$\langle H \rangle = \frac{3k}{2} \sum_{q, AA'} \text{Re}(J_q^{AA'} \mathcal{F}_q^{AA'}) + \sum_{L=2}^{\infty} \sum_{B_1 \dots B_L} \hat{u}_{B_1 \dots B_L}^{(L)} \sum_{m_1 \dots m_L} (\det \Delta^{(L-1)})^{-3/2}. \quad (31)$$

The static structure factor of light scattering in the nonequilibrium situation is defined in terms of the equal-time correlations

$$s(\mathbf{k}, t) \equiv \frac{1}{KM} \sum_{nn' BB'} \langle \exp\{i\mathbf{k} \cdot [\mathbf{x}_n^B(t) - \mathbf{x}_{n'}^{B'}(t)]\} \rangle. \quad (32)$$

In the Gaussian self-consistent approach the average may be written as

$$s(\mathbf{k}, t) = \frac{1}{KM} \sum_{nn' BB'} \exp\left(-\frac{\mathbf{k}^2}{2} D_{nn'}^{BB'}(t)\right). \quad (33)$$

It is useful also to consider the density of monomers that occupy the  $A$ th position in blocks,

$$\rho^A(\mathbf{r}) = \left\langle \sum_n \delta(\mathbf{x}_n^A - \mathbf{R} - \mathbf{r}) \right\rangle, \quad (34)$$

measured from the center of mass of the whole polymer  $\mathbf{R}$ . The calculation shows that this function is given by a three-dimensional Gaussian distribution normalized to  $M$  with variance  $G^A$ ,

$$G^A = \sum_{q \neq 0} \mathcal{F}_q^{AA} + \frac{1}{K} \sum_B Z^{AB} - \frac{1}{2K^2} \sum_{BB'} Z^{BB'}. \quad (35)$$

The function  $G^A$  has the simple meaning of the partial radius of gyration of  $A$ th monomers according to the formula

$$R_g^2 = \frac{1}{K} \sum_A G^A. \quad (36)$$

In conclusion, let us discuss some limiting cases. By setting  $\hat{u}_{B_1 \dots B_L}^{(L)} \equiv \hat{u}^{(L)}$  one should obtain a homopolymer. Thus, the correlation functions should coincide with those of a homopolymer  $D_{j,j'}^{\text{hom}} \equiv \frac{1}{3} \langle (\mathbf{x}_j - \mathbf{x}_{j'})^2 \rangle$  for corresponding monomer positions  $j = Km + B$ ,

$$D_{mm'}^{BB'} = D_{Km+B, Km'+B'}^{\text{hom}}. \quad (37)$$

By a straightforward but somewhat tedious calculation, one can demonstrate that in this limit Eq. (18) reduces to the appropriate equation for a homopolymer. Let us mention also that for a ring homopolymer  $\mathcal{F}_0^{AA}$  is independent of the block index  $A$  and thus the variables  $Y^A$  are trivial. This follows from formula (37) and the property that  $\langle \mathbf{x}_j^2 \rangle$  is  $j$  independent for a ring homopolymer.

Another curious limit is the heteropolymer with only one block  $M = 1$ , i.e., no periodic structure. In this limit the Fourier and spatial monomer positions coincide, since the chain indices could take only the value  $q = n = 0$ . Note that all our previous formulas will remain valid in this case for both ring and open polymer [17]. Thus, we shall have  $K(K+1)/2$  equations of motion to determine, generally speaking, a nondiagonal symmetric matrix  $\mathcal{F}_0^{AA}$ .

### III. THE INVERSION SYMMETRY AND REDUCTION OF THE DEGREES OF FREEDOM

We begin by noting that, although the spatial correlations  $D_{mm'}^{BB'}$  are real and positive, the quantities  $\mathcal{F}_q^{AA'}$  are complex as are the equations of motion (18). For practical applications this is a somewhat inconvenient circumstance. However, for a broad class of chain sequences there is an additional symmetry property of the chain that allows us to exclude some redundant degrees of freedom, and thereby make the correlation functions real. This symmetry is much more general than the Gaussian self-consistent method itself that we currently develop.

Let us denote by  $c(A)$  the monomer type of the  $A$ th element of a block. We say that the chain possesses inversion symmetry if by considering the chain in the reversed

direction we see the same sequence of monomer types after a shift of some number of positions  $L$ , called the rank of a sequence,

$$c(L-1-A) = c(A) \pmod{K}. \quad (38)$$

Thus, all observables should remain invariant under the transformation of the chain index generated by such shift of the chain reference point and inversion of the direction. Obviously, by the definition,  $L$  is determined only by the block sequence. Indeed, if we shift the chain reference point by  $s$  beads, and thus change the block sequence,  $L$  would be transformed by

$$L \rightarrow L + 2s \pmod{K}. \quad (39)$$

Thus, it is always possible to choose  $s$  so that  $L$  will be equal either to zero or  $K-1$  [18]. The rank for such a choice of  $s$  will be called canonical. The resulting symmetry properties can be obtained in the following formal way. Consider a block of the length  $K$  and construct the matrix  $\Phi$  by the following algorithm:

$$\Phi = \begin{pmatrix} 0 & 1 & \dots & K-1 \\ c(0) & c(1) & \dots & c(K-1) \\ 0 & 0 & \dots & 0 \end{pmatrix}. \quad (40)$$

Let us invert the matrix

$$\Phi^i = \begin{pmatrix} K-1 & K-2 & \dots & 0 \\ c(K-1) & c(K-2) & \dots & c(0) \\ 0 & 0 & \dots & 0 \end{pmatrix}. \quad (41)$$

Now we have two sequences,  $c(0)c(1)\dots c(K-1)$  and  $c(K-1)c(K-2)\dots c(0)$ . Let us perform cyclic right shifts of the matrix columns. We move the last column to the first position until we recover the original binary sequence. This is not always possible. Simple examples of irreducible sequences that do not possess this property are "abaabb" and "ababbabb". Now, consider the resulting matrix

$$\Phi^r = \begin{pmatrix} \pi(0) & \pi(1) & \dots & \pi(K-1) \\ c(0) & c(1) & \dots & c(K-1) \\ \mathcal{P}(0) & \mathcal{P}(1) & \dots & \mathcal{P}(K-1) \end{pmatrix}. \quad (42)$$

Here  $\pi(A)$  denotes the permutation of the original ordered sequence  $01\dots K-1$  producing the required sequence of positions within the block. We have also used the notation  $\mathcal{P}(A) = 1$  if  $\pi(A)$  is the image of  $A$  that has been moved from the last position to the first by cyclic shifts, and zero if it has just been shifted within the matrix range. Thus, by construction, the last row will contain some number of units  $L$ , the rank of the block, and  $K-L$  zeros. Now, we can write  $\Phi^r$  in more detail as

$$\Phi^r = \begin{pmatrix} L-1 & \dots & 0 & K-1 & \dots & L \\ c(L-1) & \dots & c(0) & c(K-1) & \dots & c(L) \\ \underbrace{1 \quad \dots \quad 1}_L & \underbrace{0 \quad \dots \quad 0}_{K-L} \end{pmatrix}. \quad (43)$$

Explicitly  $\pi(A)$  can be presented in the form

$$\pi(A) = \begin{cases} L-1-A, & 0 \leq A \leq L-1 \\ K-1+L-A, & L \leq A \leq K-1, \end{cases}$$

$$\mathcal{P}(A) = \begin{cases} 1, & 0 \leq A \leq L-1, \\ 0, & L \leq A \leq K-1 \end{cases} \quad (44)$$

and it possesses the property  $c(\pi(A)) = c(A)$ . The following identities hold:

$$D_{mm'}^{AA'} = D_{m'+\mathcal{P}(A), m+\mathcal{P}(A')}^{\pi(A), \pi(A')},$$

$$A, A' = 0, \dots, K-1, \quad m, m' = 0, \dots, M-1, \quad (45)$$

and for the Fourier variables this identity takes the form

$$\mathcal{F}_q^{AA'} = f_q^{\mathcal{P}(A')-\mathcal{P}(A)} \mathcal{F}_q^{\pi(A'), \pi(A)}. \quad (46)$$

*Lemma:* Identities (46) for any nonzero  $q$  yield  $K(K-1)/2$  real independent equations and thus the number of real independent degrees of freedom in  $\mathcal{F}_q^{AA'}$  does not depend on the block rank and is equal to  $K(K+1)/2$  if the inversion symmetry exists for the chain.

The proof is given in Appendix B.

It is of importance to construct a unitary transformation  $\mathcal{A}_q$  over the block indices,

$$\tilde{\mathcal{F}}_q \equiv \mathcal{A}_q^\dagger \mathcal{F}_q \mathcal{A}_q, \quad (47)$$

that would give rise to a real symmetric matrix  $\tilde{\mathcal{F}}_q$  with exactly  $K(K+1)/2$  for ( $q \neq 0$ ) nontrivial elements. In Appendix C we consider the construction of this transformation and here we shall only present the results. Denote by  $\Pi$  the permutation matrix corresponding to the permutation (44) for a canonical choice of rank. Let us consider first the sequence of canonical rank  $L = 0$ , for which

$$\Pi^{AA'} = \delta^{K-1-A-A', 0}. \quad (48)$$

Then, the transformation matrix is  $q$  independent and may be taken simply as

$$\mathcal{A} = \frac{e^{-i\pi/4}}{\sqrt{2}} (\Pi + i\mathbf{1}). \quad (49)$$

The case of the canonical rank  $L = K-1$  is more elaborate. The permutation matrix there has the structure

$$\Pi^{\alpha\alpha'} \equiv \check{\Pi}^{\alpha\alpha'} = \delta^{K-2-\alpha-\alpha', 0},$$

$$\Pi^{K-1K-1} = 1, \quad \Pi^{K-1\alpha} = \Pi^{\alpha K-1} = 0, \quad (50)$$

where  $\alpha, \alpha' = 0, \dots, K-2$ . Then the transformation may be taken in the form

$$\mathcal{A}_q^{\alpha\alpha'} = \check{\mathcal{A}}^{\alpha\alpha'}, \quad \check{\mathcal{A}} = \frac{e^{-i\pi/4}}{\sqrt{2}} (\check{\Pi} + i\mathbf{1}),$$

$$\mathcal{A}_q^{K-1K-1} = f_q^{1/2}, \quad \mathcal{A}_q^{K-1\alpha} = \mathcal{A}_q^{\alpha K-1} = 0. \quad (51)$$

These considerations collectively lay out the formal structure necessary to deal with all heteropolymers with a fixed sequence of monomers. However, to illustrate the approach we now turn to the simplest possible example, a copolymer with block of size two.

#### IV. *ab* SEQUENCE

Suppose that a block consists of just two monomer types  $a$  and  $b$ . The matrices (40)–(42) have the form

$$\Phi = \begin{pmatrix} 0 & 1 \\ a & b \\ 0 & 0 \end{pmatrix}, \quad \Phi^r = \begin{pmatrix} 0 & 1 \\ a & b \\ 1 & 0 \end{pmatrix}, \quad (52)$$

and Eq. (46) yields only one relation,

$$\mathcal{F}_q^{01} = f_q^* (\mathcal{F}_q^{01})^*. \quad (53)$$

Thus, the variable

$$\mathcal{F}_q^3 \equiv f_q^{1/2} \mathcal{F}_q^{01} \quad (54)$$

is real  $(\mathcal{F}_q^3)^* = \mathcal{F}_q^3$ , and the diagonal elements are real by definition. Let us perform the unitary transformation over the block indices,

$$\tilde{\mathcal{F}}_q \equiv A_q^{-1} \mathcal{F}_q A_q, \quad \tilde{J}_q \equiv A_q^{-1} J_q^T A_q, \quad (55)$$

$$\tilde{d}_{\dots}^{(q)} \equiv A_q^{-1} (d_{\dots}^{(q)})^T A_q,$$

where the transformation matrix is [19]

$$A_q = -A_q^{-1} = \begin{pmatrix} 0 & \exp(-i\pi q/2M) \\ -\exp(i\pi q/2M) & 0 \end{pmatrix}. \quad (56)$$

The resulting matrices  $\tilde{\mathcal{F}}_q$  and  $\tilde{J}_q$  are real and symmetric and have the form

$$\tilde{\mathcal{F}}_q = \begin{pmatrix} \mathcal{F}_q^{11} & -\mathcal{F}_q^3 \\ -\mathcal{F}_q^3 & \mathcal{F}_q^{00} \end{pmatrix}, \quad (57)$$

$$\tilde{J}_q = \begin{pmatrix} 2 & 2 \cos(\pi q/M) \\ 2 \cos(\pi q/M) & 2 \end{pmatrix}.$$

The equation of motion (18) now becomes explicitly real and may be written in block matrix notation as

$$\zeta \tilde{\mathcal{F}}_q = 2k_B T \mathbf{1} - k (\tilde{\mathcal{F}}_q \tilde{J}_q + \tilde{J}_q \tilde{\mathcal{F}}_q) + \sum_{L=2}^{\infty} \sum_{B_1 \dots B_L} \hat{u}_{B_1 \dots B_L}^{(L)} \sum_{m_1 \dots m_L} (\det \Delta^{(L-1)})^{-5/2} \times \sum_{i,j=1}^{L-1} \tilde{\Delta}_{ij}^{(L-1)} (\tilde{\mathcal{F}}_q \hat{d}_{ij}^{(q)} + \hat{d}_{ij}^{(q)\dagger} \tilde{\mathcal{F}}_q), \quad (58)$$

where  $\hat{d}_{ij}^{(q)} = \text{Re} \tilde{d}_{ij}^{(q)}$ . The two-body coefficients can be presented by the following matrix in indices  $AA'$ :

$$\begin{aligned} \hat{d}_{mm'BB'}^{(q)} &\equiv \text{Re} \hat{d}_{mm',mm',BB',BB'}^{(q)} \\ &= \begin{pmatrix} g_{mm'BB'}^{1(q)} & -g_{mm'BB'}^{3(q)} \\ g_{mm'BB'}^{3(q)} & g_{mm'BB'}^{0(q)} \end{pmatrix}. \end{aligned} \quad (59)$$

Here the coefficients  $g^{(q)}$  also may be written as the matrices in  $BB'$ ,

$$g_{mm'}^{0(q)} = \begin{pmatrix} 4 \sin^2 \frac{\pi q(m-m')}{M} & 1 \\ 1 & 0 \end{pmatrix}, \quad (60)$$

$$g_{mm'}^{1(q)} = \begin{pmatrix} 0 & 1 \\ 1 & 4 \sin^2 \frac{\pi q(m-m')}{M} \end{pmatrix},$$

$$g_{mm'}^{3(q)} = \begin{pmatrix} 0 & -\cos \frac{2\pi q(m-m'-1/2)}{M} \\ -\cos \frac{2\pi q(m-m'+1/2)}{M} & 0 \end{pmatrix}. \quad (61)$$

The matrix  $D_{mm'}^{BB'}$  is related to  $\tilde{\mathcal{F}}_q^{AA'}$  by

$$D_{mm'}^{BB'} = \sum_{q=1}^{M-1} \text{tr}(\hat{d}_{mm'BB'}^{(q)} \tilde{\mathcal{F}}_q) + Z(\delta^{B0} \delta^{B'1} + \delta^{B1} \delta^{B'0}), \quad (62)$$

and the radius of gyration is equal to

$$R_g^2 = \frac{1}{2} \sum_{q=1}^{M-1} (\mathcal{F}_q^{00} + \mathcal{F}_q^{11}) + \frac{Z}{4}. \quad (63)$$

Note that we have separated the contribution of the zero mode  $Z$  from the rest of the expression. We can define three independent linear combinations of the zero modes according to (27)–(30),

$$W = \mathcal{F}_0^{00} + \mathcal{F}_0^{11} + 2\mathcal{F}_0^3, \quad (64)$$

$$Y \equiv Y^1 = \mathcal{F}_0^{00} - \mathcal{F}_0^{11}, \quad Z \equiv Z^{01} = \mathcal{F}_0^{00} + \mathcal{F}_0^{11} - 2\mathcal{F}_0^3. \quad (65)$$

These variables satisfy the equations of motion,

$$\zeta \dot{W} = 4k_B T, \quad (66)$$

$$\zeta \dot{Y} = 2(-2k + I^{(0)})Y, \quad (67)$$

$$\zeta \dot{Z} = 4k_B T + 4(-2k + I^{(0)})Z, \quad (68)$$

where  $I^{(0)}$  is given by the relation

$$\begin{aligned} I^{(0)}(Z) &= \sum_{L=2}^{\infty} \sum_{B_1 \dots B_L} \hat{u}_{B_1 \dots B_L}^{(L)} \sum_{m_1 \dots m_L} (\det \Delta^{(L-1)})^{-5/2} \\ &\times \sum_{i,j=1}^{L-1} \bar{\Delta}_{ij}^{(L-1)} \hat{d}_{ij}^{(0)00}. \end{aligned} \quad (69)$$

Equation (66) describes the diffusion of the center of mass of the whole polymer. Equation (67) is homogeneous and thus, for our problem, it has only a trivial solution  $Y = 0$ , this being a consequence of a homopolymer initial condition. Moreover, even for a nontrivial initial condition from Eqs. (67) and (68) one can see that  $Y(t) \rightarrow 0$  at  $t \rightarrow \infty$ . Finally, Eq. (68) describes those conformational motions that affect the mean square difference between the centers of mass of  $a$  and  $b$  monomers. Only the variable  $Z$  contributes to (62) and (63) and it may be considered as an order parameter characterizing the global spatial distribution of monomers of both types. As we shall see, it is one of the good order parameters in describing the kinetics of collapse of the  $ab$  copolymer.

## V. NUMERICAL RESULTS

In this section we present the results from numerical solution of the self-consistency equations (58) for the  $ab$  sequence. Our main purpose here is to illustrate the advantages of the Gaussian self-consistent approach for an example of equilibrium and kinetics phenomena. However, being the simplest example of this type of situation, this study uncovers an interesting physical phenomenon—essentially a type of phase separation accompanying the coil-to-globule transition.

It is known [10,20] that for most cases it is sufficient to account for the excluded volume effect only up to the three-body interactions, i.e., one may set  $u^{(L)} = 0$  for  $L > 3$ . Moreover, one may assume that the third virial coefficients are the same for both types of monomers  $u_{BB'B''}^{(3)} \equiv u^{(3)}$ , so the only difference between  $a$  and  $b$  monomers will be in the second virial coefficients. Let us introduce the following useful parametrization of the second virial coefficients,

$$\bar{u}^{(2)} = \frac{1}{4}(u_{00}^{(2)} + u_{11}^{(2)} + 2u_{01}^{(2)}), \quad (70)$$

$$u_e^{(2)} = \frac{1}{2}(u_{00}^{(2)} - u_{11}^{(2)}), \quad (71)$$

$$u_d^{(2)} = \frac{1}{4}(u_{00}^{(2)} + u_{11}^{(2)} - 2u_{01}^{(2)}). \quad (72)$$

There are several interesting physical problems corresponding to different planes in the three-dimensional phase space. We may mention the much studied so-called “charge” model [21,22], where  $u_e^{(2)} = 0$ . Here we shall consider a less analytically tractable but more interesting case from the point of view of applications to biopolymers,  $u_d^{(2)} = 0$ . It arises naturally in the lattice model of copolymers [13], if one assumes that  $a$  and  $b$  monomers have equation interaction constants with each other and only differ by their interaction constants with the solvent. In Ref. [23] the condition  $u_d^{(2)} = 0$  has been also chosen as it is related to the incompressibility condition of the solvent. Let us assume that  $u_{00}^{(2)} \leq u_{11}^{(2)}$ , so that  $a$  monomers might be interpreted as hydrophobic and  $b$  as hydrophilic. It is natural to work with the combinations  $\mathcal{L} = (k_B T / \kappa)^{1/2}$  and  $\mathcal{T} = \zeta_b / \kappa$  as the units of

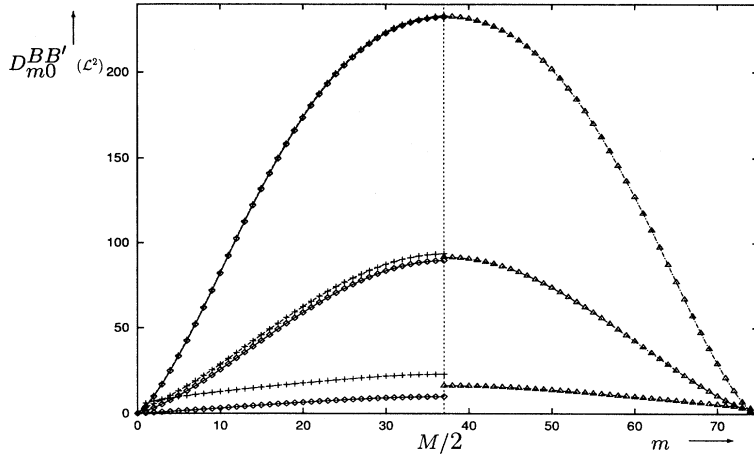


FIG. 1. Plots of the equilibrium values of the space correlations  $D_{m0}^{00}$  (diamonds),  $D_{m0}^{11}$  (crosses), and  $D_{m0}^{01}$  (triangles) vs the chain index  $m$  for different values of the second virial coefficient  $\bar{u}^{(2)}$  (from top to bottom):  $\bar{u}^{(2)} = 5$ ,  $\bar{u}_c^{(2)}$ , and  $-25$ . For convenience we have drawn the plots only on one of the sides, and they may be extended to the other one by the inversion symmetry properties.

size and time in the system. In the sequel we have used the following particular choice of parameters:  $k_B T = 1$ ,  $\kappa = 1$ , and  $\zeta_b = 1$ , which fix  $\mathcal{L}$  and  $\mathcal{T}$  equal to unity.

Here we are interested in studying the effect of the hydrophobicity at equilibrium and in kinetics. Thus, we work with a fixed second virial coefficient of hydrophilic species. In our numerical study we have fixed this as  $u_{11}^{(2)} = 15k_B T \mathcal{L}^3$ , and also  $u^{(3)} = 10k_B T \mathcal{L}^6$  [24]. Hence, we shall consider only a line in the phase space of the second virial coefficients parametrized by only one variable  $\bar{u}^{(2)}$ . This variable has the simple meaning of the “mean” second virial coefficient between both species. The point  $\bar{u}^{(2)} = u_{11}^{(2)}$  corresponds exactly to the homopolymer in the Flory coil.

### A. Equilibrium

Setting the time derivative in (58) to zero, one gets a set of nonlinear algebraic equations for determination of the equal-time correlation functions at equilibrium for a given value of  $\bar{u}^{(2)}$ .

The region  $\bar{u}^{(2)} \lesssim u_{11}^{(2)}$  corresponds to the good solvent condition. We find that in this region all the observables we considered are very close to those of a homopolymer with the second virial coefficient  $\bar{u}^{(2)}$ . Indeed, the overall excluded volume interaction is repulsive here, and its strength is proportional to the mean second virial coefficient between beads, i.e.,  $\bar{u}^{(2)}$ . The upper curves in Fig. 1 show that the mean squared distances between like monomers along the chain are very close to each other in this region,  $D_{m0}^{00} \simeq D_{m0}^{11}$ . One can show that using this equality Eq. (58) becomes the equation for a homopolymer with  $u^{(2)} = \bar{u}^{(2)}$ . Thus, the mean-field theory for this region may be conceived as the theory of an effective homopolymer.

This approximation breaks down near the  $\theta$  point  $\bar{u}^{(2)} \simeq 0$ . In Fig. 2 we exhibit the behavior of the squared radius of gyration,  $R_g^2$ , the heat capacity,  $C_V \equiv (1/2)\partial\langle H \rangle / \partial \bar{u}^{(2)}$ , and the density difference of hydrophobic and hydrophilic species in the center of mass of the whole polymer,  $\Delta\rho \equiv \rho^0(\mathbf{0}) - \rho^1(\mathbf{0})$ , in the vicinity of this point. At  $\bar{u}^{(2)} = \bar{u}_c^{(2)}$  the heat capacity has a sharp peak. This may be a signature of a first-order-like transition. On the one hand, the radius of gyration falls strongly

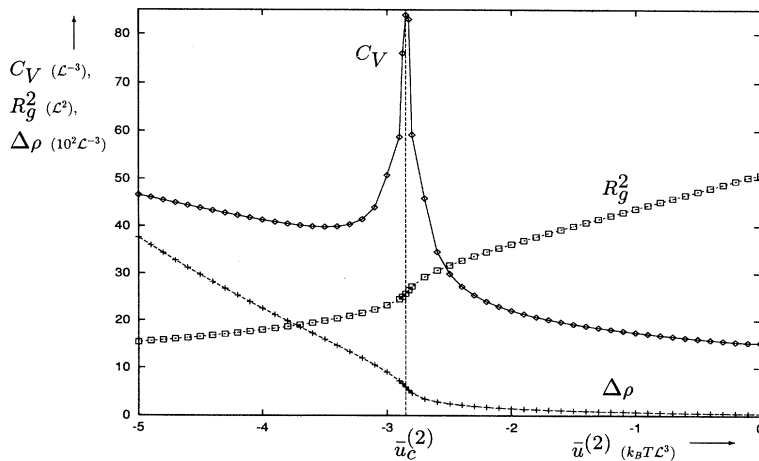


FIG. 2. Equilibrium plots of the squared radius of gyration  $R_g^2$ , the heat capacity  $C_V$ , and the density difference of hydrophobic and hydrophilic species  $\Delta\rho$  vs the second virial coefficient  $\bar{u}^{(2)}$  across the transition. The point  $\bar{u}_c^{(2)}$  corresponds to the critical point, in which the heat capacity reaches its maximum.

TABLE I. Values of the equilibrium radii of gyration vs the degree of polymerization  $N$  for the heteropolymer ( $u_e^{(2)} = -40$ ) and the homopolymer ( $u_e^{(2)} = 0$ ) collapsed globules. Here  $\nu$  denotes the swelling exponent of the scaling law  $R_g \sim N^\nu$ .

Radius of gyration	$N$						$\nu$
	30	50	70	100	150	200	
$R_g^2(\bar{u}^{(2)} = -25, u_e^{(2)} = 0)$	1.186	1.910	2.585	3.528	4.956	6.247	$0.42 \pm 0.03$
$R_g^2(\bar{u}^{(2)} = -25, u_e^{(2)} = -40)$	1.858	2.461	3.024	3.878	5.513	7.502	$0.36 \pm 0.04$

near this point, similar to the ordinary coil-to-globule transition. We note that there is a new order parameter, namely,  $\Delta\rho$ , that almost vanishes for  $\bar{u}^{(2)} > \bar{u}_c^{(2)}$  and grows quickly in the poor solvent region.

The latter parameter reflects the degree of microphase separation of  $a$  and  $b$  components. The phase separation leads to the equilibrium globule where the hydrophobic species tend to be in the core of the globule and the hydrophilic ones form the exterior. Of course, the Gaussian self-consistent approach imposes special restrictions on the density distribution functions  $\rho^A(\mathbf{r})$ , since these can only be a Gaussian shape. Thus, the phase separation in our method leads to a wider dispersion of hydrophilic beads,  $G^1$ , compared to the dispersion of hydrophobic ones,  $G^0$ . One can see this from Fig. 2 and the relation

$$\Delta\rho = (2\pi)^{-3/2}[(G^0)^{-3/2} - (G^1)^{-3/2}]. \quad (73)$$

Similarly, the phase separation manifests itself in the splitting between the functions  $D_{m0}^{00}$  and  $D_{m0}^{11}$  in Fig. 1. The gap is distinguishable but relatively small at the critical point (middle curves), and is quite pronounced for large negative values of the second virial coefficient (lower curves).

Concerning the order of the phase transition, it obviously requires a careful analysis of the system with respect to size scalings. This analysis has not as yet been accomplished by us for sufficiently long chains. Our current thinking is that this transition grows to become discontinuous (first order) in the infinite size limit but we emphasize the novel feature of the transition itself. Thus, not only is the fractal dimension of the system changing, but there is, as we have seen, phase separation of

the hydrophobic and hydrophilic units. Indeed we also comment that it is noteworthy that there appears to be only one transition rather than consecutive ones for the different processes. In any case, these are rather subtle questions that would require a separate and rather comprehensive study, and our emphasis here lies on the generic kinetics features of the problem.

Finally, we would like to compare the radii of gyration of the heteropolymer and the homopolymer collapsed globules. The results are presented in Table I for the same value of  $\bar{u}^{(2)}$  and two different values of  $u_e^{(2)}$ . From those data one can find that for sufficiently long chains the scaling law  $R_g^2 = AN^{2\nu}$  is valid for both cases with  $\nu = \frac{1}{3}$  with the constant  $A$  being larger for the heteropolymer. This may be explained by the observation that the heteropolymer globule possesses a less dense hydrophilic exterior compared to the completely compacted homopolymer globule.

## B. Kinetics

Now let us consider the kinetics after a quench  $\bar{u}^{(2)} = u_{11}^{(2)} \rightarrow \bar{u}^{(2)'}$  from the homopolymer Flory coil to the region  $\bar{u}^{(2)' < \bar{u}_c^{(2)}$ . We refer the reader to a previous work [10] for the details of the homopolymer kinetics.

It turns out that for some time after this quench the behavior of the global observables of the system is close to that for an effective homopolymer that has undergone the same quench. In Fig. 3 we exhibit the complete,  $R_g^2$ , and partial,  $G^A$ , squared radii of gyration, such that  $R_g^3 = (G^0 + G^1)/2$ , as well as the squared radius of gy-

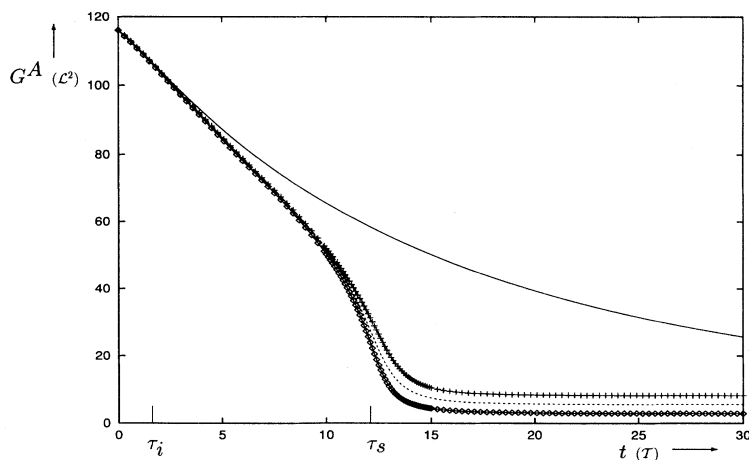


FIG. 3. Plots of the partial,  $G^0$  (diamonds),  $G^1$  (crosses), and the complete,  $R_g^2$  (dashed line), radii of gyration vs time  $t$  for polymer with the degree of polymerization  $N = 150$ . The final value of the second virial coefficient is equal to  $\bar{u}^{(2)' = -25$ . The solid line represents the squared radius of gyration of homopolymer after the same quench. The value  $\tau_i$  has been taken from the results of Ref. [10]. Point  $\tau_s$  corresponds to the characteristic time of the phase separation.



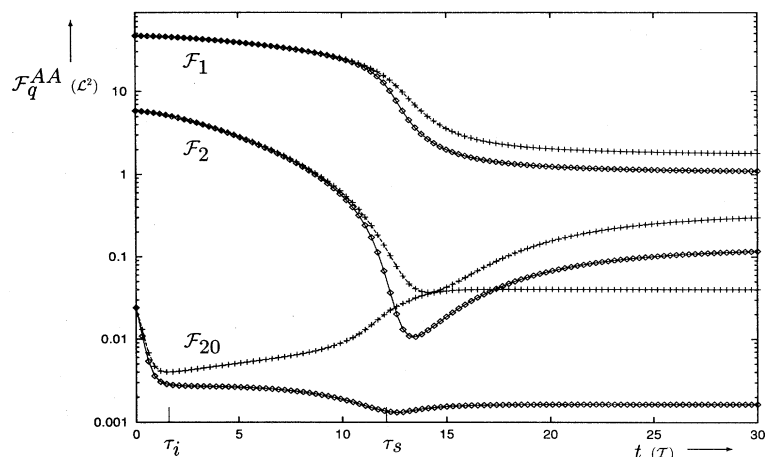


FIG. 4. Plots of the Fourier modes  $\mathcal{F}_q^{00}$  (diamonds) and  $\mathcal{F}_q^{11}$  (crosses) vs time  $t$  for polymer with the degree of polymerization  $N = 150$  at different values of the chain index  $q$  (from top to bottom):  $q = 1, 2,$  and  $20$ . The final value of the second virial coefficient is equal to  $\bar{u}^{(2)'} = -25$ .

ration of the effective homopolymer. They coincide approximately up to the time  $\tau_i$ , which we have previously interpreted as the duration of the first (spinodal) kinetic stage of homopolymer [9,10]. Thus, for sufficiently small quenches the same kinetic law is valid during this stage,

$$R_q^2(t) = R_q^2(0) - At^{\alpha_i}, \quad t < \tau_i, \quad (74)$$

where  $\alpha_i = (2 - \nu_F)/(2\nu_F + 1)$  with  $\nu_F = \frac{3}{5}$  being the Flory exponent. The duration of this stage,  $\tau_i$ , is essentially independent of the degree of polymerization, reflecting the localized character of cluster growth mechanism at this stage. However, from Fig. 4, where we draw the time evolution of the internal modes  $\mathcal{F}_q^{AA}(t)$ , one can see that large- $q$  modes describing conformations on small distances along the chain (lower curves) already show certain deviation from each other. This means that, although our spinodal picture of the first stage is still correct, the local structure of the clusters that are forming is different from those in the homopolymer. Visualization of polymer conformations in Monte Carlo simulation of the system [13,25] shows that these clusters look

like small prototypes of the copolymer collapsed globule with a tendency for a hydrophilic exterior and hydrophobic interior. Meanwhile, the small- $q$  modes that describe conformations on large scales (upper curves) do not yet exhibit this splitting.

The difference between  $G^A$  and the homopolymer squared radius of gyration for  $t_i < t < \tau_s$  in Fig. 3 being significant, there is almost no difference here between  $G^0$  and  $G^1$ . In fact, all large-scale characteristics of the system are quite similar for hydrophobic and hydrophilic species up to some time  $\tau_s$ . For example, in Fig. 4 this is true for all sufficiently small- $q$  modes,  $\mathcal{F}_q^{AA}$ . Thus, the difference between contributions of  $a$  and  $b$  monomers to such global characteristics remains almost negligible up to the time  $\tau_s$ , when there is a precipitous change reflecting a sort of phase separation. We note that around this time the rapid change is driven by the small- $q$  modes.

Let us consider the time evolution of the mean squared difference between the center of mass of hydrophobic and hydrophilic monomers,  $Z$ , presented in Fig. 5. During the beginning of the spinodal stage  $Z$  drops significantly, which is natural for the high- $q$  modes of homopolymer. Then it remains nearly constant for a comparatively long

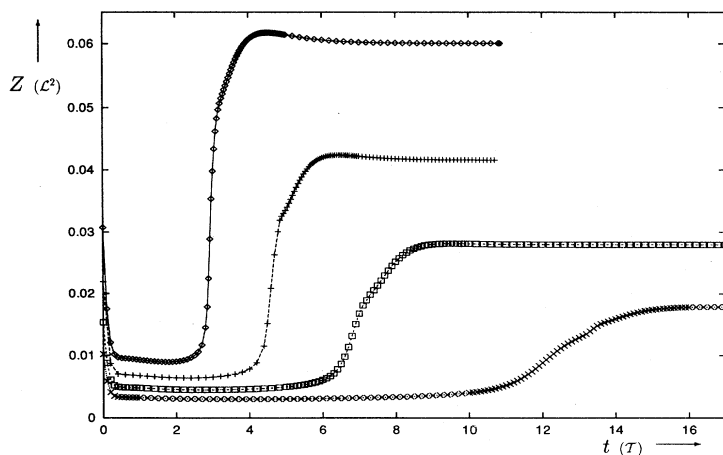


FIG. 5. Plots of the mean squared distance between the center of mass of  $a$  and  $b$  monomers  $Z$  vs time  $t$  for polymers with the degrees of polymerization (from top to bottom):  $N = 50, 70, 100,$  and  $150$ . The final value of the second virial coefficient is equal to  $\bar{u}^{(2)'} = -25$ .

TABLE II. Values of the characteristic time of the phase separation  $\tau_s$  and the final relaxation time  $\tau_f$  vs the degree of polymerization  $N$  for different second virial coefficients  $\bar{u}^{(2)'}$ . Here the quantity  $\gamma$  denotes the exponent of the appropriate time in terms of the degree of polymerization, i.e.,  $\tau \sim N^\gamma$ .

Time	$N$							$\gamma$
	20	30	50	70	100	150	200	
$\tau_s(\bar{u}^{(2)' = -4)$	5.35	12.0	30.4	53.3	87.2	166	240	$1.57 \pm 0.09$
$\tau_s(\bar{u}^{(2)' = -10)$	2.20	4.32	9.70	15.8	24.1	42.8	59.6	$1.36 \pm 0.08$
$\tau_s(\bar{u}^{(2)' = -25)$	0.80	1.44	2.93	4.60	6.85	12.1	17.0	$1.30 \pm 0.04$
$\tau_s(\bar{u}^{(2)' = -50)$	0.338	0.57	1.09	1.55	2.48	4.37	6.22	$1.25 \pm 0.03$
$\tau_f(\bar{u}^{(2)' = -4)$	1.0	2.3	5.2	12	22	44	84	$1.90 \pm 0.08$
$\tau_f(\bar{u}^{(2)' = -25)$	0.14	0.33	0.73	1.22	2.26	4.37	7.14	$1.67 \pm 0.07$

time, after which it suddenly begins to increase rapidly. Thus, practically it is simple to define  $\tau_s$  as the point of maximum of the time derivative of  $Z$ . From Fig. 3 one can see that  $\tau_s$  defined in this manner really may be considered as the characteristic time of the phase separation. It is also clear from Fig. 4 that all internal modes undergo rapid changes on a narrow time interval around  $\tau_s$ . This interplay between different observables is a signature of the strong mode coupling during the phase separation. In Table II we present the numerical values  $\tau_s$  for different degrees of polymerization and quenches. Using the results for homopolymer kinetics from Ref. [10] we can obtain the following bound on the time  $\tau_s$ :

$$\tau_i < \tau_s < \tau_m, \quad (75)$$

where  $\tau_m \sim N^2$  is the characteristic time of the middle stage of homopolymer kinetics. This bound suggests that the dependence of  $\tau_s$  on the degree of polymerization is also a power law  $\tau_m \sim N^{\gamma_s}$ , and in Table II we have presented numerical values of this exponent for different quenches. There is clearly some crossover in this exponent, which appears to be higher for smaller quenches.

The long plateau approximately between  $\tau_i$  and  $\tau_s$  in  $Z$  may be considered some kind of metastability in kinetics. To be able to make such an interpretation we should

study the time evolution of the mean energy (31), and this is exhibited in Fig. 6. Indeed, the mean energy falls strongly during the first kinetic stage. Then, the plateau in  $Z$  corresponds exactly to the region of slower decrease in the mean energy, which is terminated by a new rapidly falling regime at  $\tau_s$ . Metastability usually arises due to some potential barrier. Thus, after the first stage there is a chain of locally collapsed clusters that have already phase separated. The hydrophilic shells on the exterior of these clusters partially screen the attractive interaction between their cores and impede their further unification. A remaining overall attraction still draws these clusters together, and they form a crumpled low density globule as the overall radius continues to decrease. Thus, there is no true metastable state for the system studied, but it is clear that the number of available directions in the coordinate space that the system can move in is greatly reduced because barriers have been created.

Note that if the excluded volume effects (both attraction and repulsion) were of a finite scale and the block length longer, it would be possible to achieve almost a complete screening of the hydrophobic attraction by the hydrophilic repulsive shells. In that case the metastability would become profound (e.g., due to a random distribution such as in Ref. [13]), and we would see a very long metastability time analogous to the spin glass kinet-

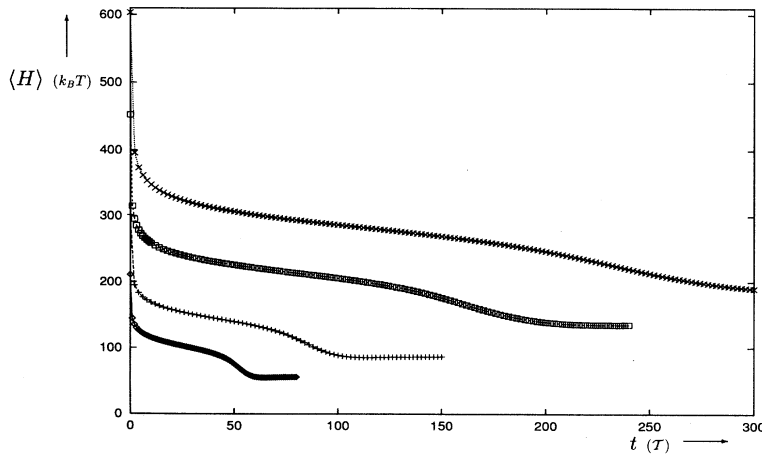


FIG. 6. Plots of the mean energy  $\langle H \rangle$  vs time  $t$  for polymers with the degrees of polymerization (from bottom to top):  $N = 70, 100, 150, \text{ and } 200$ . The final value of the second virial coefficient is equal to  $\bar{u}^{(2)' = -4}$ .

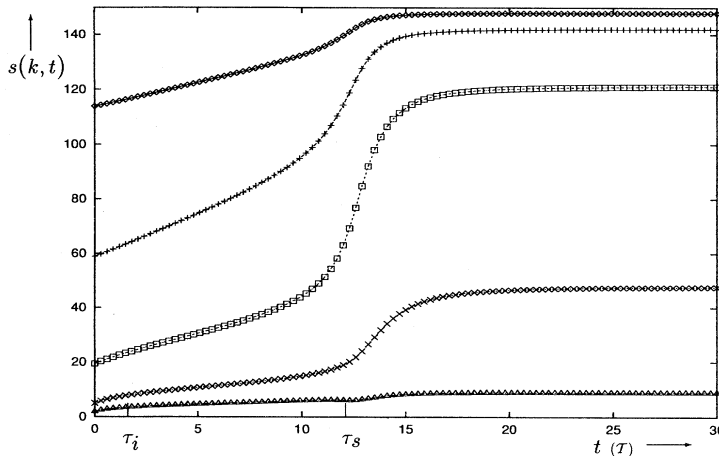


FIG. 7. Plots of the nonequilibrium static structure factor  $s(k, t)$  vs time  $t$  for polymer with the degree of polymerization  $N = 150$  at different wave numbers  $k$  (from top to bottom):  $k = 0.05, 0.1, 0.2, 0.5,$  and  $1.0$ . The final value of the second virial coefficient is equal to  $\bar{u}^{(2)'} = -25$ .

ics [26].

Now, our globule continues to undergo a uniform shrinking due to effective attraction, and when it becomes relatively dense, the repulsive barrier of the hydrophilic shells is overcome. There appear certain collective excitations of the system that repel the hydrophilic components to the surface of the globule. The more hydrophilic units that are removed to the surface, the stronger the overall attraction becomes. Thus, the globule experiences an abrupt contraction accompanied by the simultaneous global phase separation of the densities. We deal here with a strong nonlinear collective effect, which manifests itself in such nonregular time dependencies as we have seen. The process stops when the shrinking force becomes balanced with the excluded volume effect.

Now we should address a seeming paradox in Fig. 3. Indeed, it appears strange that despite the metastability in the behavior of the time evolution of the copolymer, its collapse proceeds much faster during the coarsening stage prior to the phase separation than that of the effective homopolymer (solid line). In fact, the concept of such an effective homopolymer is only useful when there is effective averaging of the second virial coefficients between both components. This happens while the distribution is, in the ensemble average sense, uniform along the chain. As we know, a small-scale phase separation during the earliest kinetic stage violates such uniformity at about the end of this stage. Therefore after the first kinetic stage the rate of collapse is determined by  $u_{00}^{(2)}$  rather than by  $\bar{u}^{(2)}$ . Since the middle stage characteristic time scales for homopolymer [10] at  $\tau_m \sim |u^{(2)}|^{-2}$  and in our case the quench is deeper  $|u_{00}^{(2)}| > |\bar{u}^{(2)}|$ , our copolymer collapses much faster. We emphasize therefore that caution should be used in applying "smeared" mean-field approximations to discuss heteropolymers, although such ideas are certainly attractive due to their relative simplicity.

Let us also mention that the global phase separation effect should be directly observable experimentally. For example, in Fig. 7 we present the nonequilibrium static

structure factor (32) as a function of time for different wave numbers  $k$ . This quantity also grows strongly near  $\tau_s$ , and this effect is especially pronounced for moderate  $k$ , which probes the internal structure of the polymer.

Finally, there is a relaxational stage describing further compaction and shape optimization of the globule. During this stage all global characteristics change relatively little. Their deviation from the final equilibrium state is a single exponential. The final relaxation time  $\tau_f$  of the square radius of gyration is presented in Table II. The dependence of this time on the degree of polymerization appears to be similar to the homopolymer [10],  $\tau_f \sim N^{5/3}$ .

## VI. CONCLUSION

The method we have outlined is the natural extension of the homopolymer work [9,10,15] that has been quite successful so far. It has been possible to reduce the equations to a tractable numerical form, though of course a complicated block structure leads to quite a large set of equations.

We have numerically studied in much detail the simplest case of an  $ab$  sequence polymer at equilibrium and the kinetics on a line in the phase space. At equilibrium we have observed a first-order-like transition from the homopolymer Flory coil to the phase separated globule, with a hydrophilic exterior and mainly hydrophobic core. Kinetics after the heteropolymer quench proceeds in the same way as the homopolymer kinetics at the earliest stage and has a certain similarity to it during most of the coarsening stage. However, as the polymer approaches close to a single globule there appears to be a new kinetic stage after a time  $\tau_s$  describing the large-scale phase separation of hydrophilic and hydrophobic units. This is characterized by a rapid change of all observables and is, in fact, a nonlinear collective phenomenon arising due to strong mode coupling. The final stage of collapse kinetics is also analogous to homopolymer, being a single exponential relaxational process towards the final equi-

librium state.

Thus, in the case we have studied the three stages of homopolymer are essentially preserved, with the same laws being applicable excluding the middle stage. A further stage is inserted into the process after a characteristic time  $\tau_s$  that scales with a power of  $N$ . Thus, unlike conventional phase separation the whole polymer chain is involved in the separation of hydrophobic and hydrophilic units. The process is seen to be driven by the collective (small- $q$ ) modes, and naturally this is reflected in the, at first surprising, conclusion that the time of onset of rapid change in a quasi-first-order transition should depend on the large number  $N$ .

Here we would like to discuss some of the features of the present formalism that should remain of importance to applications of the method to heteropolymers with more complicated block structure. Clearly, there are no serious obstacles to their study apart, of course, from purely computational restrictions as the number of dynamical variables grows roughly as  $MK^2$ . Thus, to be able to make significant progress for the whole range of chain compositions one has to use an additional ansatz that in one way or another takes into account additional structural properties of a block. For example, such properties are expected for a block consisting of two homogeneous subblocks of lengths  $K'$  and  $K''$  each composed of the same units. This simplification of the equations would allow us to study different "thermodynamic" limits: (1)  $M \rightarrow \infty$ ,  $K = \text{const}$ ; (2)  $K \rightarrow \infty$ ,  $M = \text{const}$ ; (3)  $M \rightarrow \infty$ ,  $K \rightarrow \infty$ ,  $M/K = \text{const}$ . Some of the equilibrium and kinetic properties of the phase transition may in principle depend on which of these limits is used.

On the one hand, for arbitrary blocks, we shall have the same order parameters,  $G^A$  and  $Z^{AA'}$ . From preliminary Monte Carlo simulations [13,25] we have also seen that the basic phenomenon remains essentially the same, with modifications only in the concrete laws. A better screening of the hydrophobic cores for longer blocks gives rise to a stronger metastability in kinetics, which may even appear in such quantities as the radius of gyration. One may expect this to occur for the following reasons. The variables  $Z^{AA'}$ , since they directly reflect the degree of the phase separation between different components, should qualitatively behave in the same manner as  $Z$  for  $ab$  sequence. From Eq. (26) it is also clear that the contribution of these modes becomes more significant with increasing of the block length  $K$  or, equivalently, decreasing of the number of blocks  $M$ . In the limit  $M = 1$  only the variables  $Z^{AA'}$  contribute to the radius of gyration.

Thus, the metastable plateau of  $Z^{AA'}$  may manifest itself in the radius of gyration as well.

In summation then, we find that the inclusion of amphiphility into the model, and the weak form of frustration thereby implied, introduces a new kinetic effect that may lead to an arrest of the folding process, followed after some time by a rapid collective separation of the different types of units. We comment on the accompanying laws, but there remains some uncertainty there. Nevertheless, we believe that the phenomenon is generic, and we would expect to see something of this kind of phenomenon in experiments.

Finally, having given some attention to the formalities of the method, we have laid the foundations for a general understanding of folding of periodic heteropolymers. We believe that, as a consequence, the way is now open to permit extraction of any generic processes that govern folding, and their accompanying physical laws of kinetics. It cannot be long before this is accomplished for periodic heteropolymers, and, though there are some remaining complications for random systems, we have recently been able to show [27] that there are now no fundamental barriers to a direct attack on that problem too. If this is the case and indeed, as we currently believe, there are universal kinetic laws for these systems, then they may soon lie within the grasp of theory.

#### ACKNOWLEDGMENTS

The authors acknowledge interesting discussions with Professor A. Yu. Grosberg, Professor A. R. Khokhlov, Professor P. Pincus, Professor Y. Rabin, Professor K. Yoshikawa, and Dr. N. A. Sveshnikov, and helpful assistance from P. Kiernan and V. Kosmodamiansky. This work was supported by DEC and the Irish Government.

#### APPENDIX A: INCLUSION OF HYDRODYNAMICS

Let us briefly discuss the treatment of the Oseen tensor in our method. Thus, Eq. (10) will be replaced by

$$\dot{x}_q^{\alpha B} = \sum_{\alpha' B' q'} H_{qq'}^{\alpha\alpha' BB'} \phi_q^{\alpha' B'} , \quad (\text{A1})$$

where  $\phi_q^{\alpha' B'}$  is the right-hand side of Eq. (10) and the Oseen tensor is

$$H_{qq'}^{\alpha\alpha' BB'} = \frac{1}{\eta_s M^2} \sum_{nn'} f_n^{(q)} f_{n'}^{(-q')} \int \frac{d\mathbf{w}}{(2\pi)^3} \frac{\mathcal{P}_{\alpha\alpha'}(\hat{\mathbf{w}})}{w^2} \exp \left( -i\mathbf{w} \sum_{pA} c_{nn'BB'}^{(p)A} \mathbf{x}_p^A(t) \right) + \frac{1}{M\zeta_b} \delta^{\alpha\alpha'} \delta_{qq'} \delta^{BB'} , \quad (\text{A2})$$

$$\mathcal{P}_{\alpha\alpha'}(\hat{\mathbf{w}}) \equiv \delta_{\alpha\alpha'} - \hat{w}_\alpha \hat{w}_{\alpha'} , \quad \hat{w}_\alpha = w_\alpha / w ,$$

where  $(n, B) \neq (n', B')$  and  $\eta_s$  denotes the viscosity of the solvent. Calculating its average we obtain the inverse mobility matrix,

$$\langle H_{qq'}^{\alpha\alpha' BB'} \rangle = \delta^{\alpha\alpha'} \delta_{qq'} (\zeta_q^{-1})^{BB'} . \quad (\text{A3})$$

A simple deviation yields the result

$$(\zeta_q^{-1})^{BB'} = \frac{\delta^{BB'}}{M\zeta_b} + v_1 \sum_{nn'} \frac{\exp[i2\pi q(n-n')/M]}{(D_{nn'}^{BB'})^{1/2}} , \quad (\text{A4})$$

where  $v_1 = 1/[3(2\pi^3)^{1/2}\eta_s M^2]$ . Consequently in the ‘‘preaveraged’’ approximation [2,15] Eq. (18) can be replaced by

$$\dot{\mathcal{F}}_q^{A'A} = \sum_{B'} [(\zeta_q^{-1})^{AB'} \Gamma_q^{A'B'} + (\zeta_q^{-1})^{A'B'} \Gamma_{-q}^{AB'}] , \quad (\text{A5})$$

where  $\Gamma_q^{A'A}(t) \equiv \frac{1}{3} \langle \mathbf{x}_{-q}^{A'}(t) \phi_q^A(t) \rangle$ , which has been calculated earlier,

$$\Gamma_q^{A'A} = k_B T \delta^{A'A} - k \sum_{A''} J_q^{AA''} \mathcal{F}_q^{A'A''} + \sum_{L=2}^{\infty} \sum_{B_1 \dots B_L} \hat{u}_{B_1 \dots B_L}^{(L)} \sum_{m_1 \dots m_L} (\det \Delta^{(L-1)})^{-5/2} \sum_{i,j=1}^{L-1} \bar{\Delta}_{ij}^{(L-1)} \sum_{A''} d_{ij}^{(q)AA''} \mathcal{F}_q^{A'A''} . \quad (\text{A6})$$

After the unitary transformation to real variables (A5) will become

$$\dot{\tilde{\mathcal{F}}}_q = \tilde{\Gamma}_q \tilde{\xi}_q + \tilde{\xi}_q \tilde{\Gamma}_q , \quad (\text{A7})$$

where  $\tilde{\Gamma}_q = \mathcal{A}_q^{-1} \Gamma_q \mathcal{A}_q$  and  $\tilde{\xi}_q = \mathcal{A}_q^{-1} (\zeta_q^{-1})^T \mathcal{A}_q$ . For the  $ab$  sequence of Sec. IV we obtain after the transformation

$$\tilde{\xi}_q = \begin{pmatrix} \xi_q^{11} & -\xi_q^{01} \\ -\xi_q^{10} & \xi_q^{00} \end{pmatrix} , \quad (\text{A8})$$

with  $\xi_q^{10} = \xi_q^{01}$ . These mobility coefficients are given by

$$\xi_q^{BB'} = \frac{\delta^{BB'}}{M\zeta_b} + v_1 \sum_{nn'} \frac{e_{nn'}^{(q),BB'}}{(D_{nn'}^{BB'})^{1/2}} , \quad (\text{A9})$$

where we used the notation

$$e_{nn'}^{(q),00} = e_{nn'}^{(q),11} = \cos \frac{2\pi q(n-n')}{M} , \quad (\text{A10})$$

$$e_{nn'}^{(q),01} = e_{n'n}^{(q),10} = \cos \frac{2\pi q(n-n'-1/2)}{M} . \quad (\text{A11})$$

## APPENDIX B: PROOF OF THE LEMMA

Indeed, let us consider independently subsets of  $L$  and  $K-L$  block indices corresponding to  $\mathcal{P} = 1$  and 0. Then, when both  $A, A'$  belong to the first subset, Eqs. (46) may be written as

$$\mathcal{F}_q^{A,A'} = \mathcal{F}_q^{L-1-A',L-1-A} , \quad 0 \leq A \leq A' \leq L-1 , \quad (\text{B1})$$

and the number of real independent equations is equal to

$$n_{LL,\text{diag}} = \begin{cases} L/2 , & L \text{ even} \\ (L-1)/2 , & L \text{ odd} , \end{cases} \quad (\text{B2})$$

$$n_{LL,\text{nondiag}} = \begin{cases} L(L-2)/4 , & L \text{ even} \\ (L-1)^2/4 , & L \text{ odd} , \end{cases} \quad (\text{B3})$$

$$n_{LL} \equiv n_{LL,\text{diag}} + 2n_{LL,\text{nondiag}} = L(L-1)/2 . \quad (\text{B4})$$

Similarly for another subset we have

$$n_{K-L,K-L} = (K-L)(K-L-1)/2 . \quad (\text{B6})$$

Also there is the mixed case for which we can write Eqs. (46) as

$$\mathcal{F}_q^{A,L+A'} = f_q^* \mathcal{F}_q^{K-1-A',L-1-A} , \quad 0 \leq A \leq A' \leq L-1 , \quad (\text{B7})$$

and the number of real independent equations is

$$n_{L,K-L} = L(K-L) . \quad (\text{B8})$$

Summing up we finally obtain  $n_K$  independent real equations

$$n_K \equiv n_{L,L} + n_{K-L,K-L} + n_{L,K-L} = K(K-1)/2 . \quad (\text{B9})$$

## APPENDIX C: CONSTRUCTION OF THE UNITARY TRANSFORMATION

Let us consider identity (46) and formula (44) separately for  $L = 0$  and  $L = K-1$ .

(1) In the  $L = 0$  case formula (44) becomes

$$\pi(A) = K-1-A , \quad \mathcal{P}(A) = 0 , \quad 0 \leq A \leq K-1 , \quad (\text{C1})$$

and hence we can write (46) in the matrix form

$$\mathcal{F}_q = \Pi \mathcal{F}_q^T \Pi , \quad (\text{C2})$$

assuming notation (48). Thus, if we find  $\mathcal{A}$  a square root of  $\Pi$ , such that

$$\mathcal{A}^2 = \Pi, \quad \mathcal{A}^T = \mathcal{A}, \quad \mathcal{A}^\dagger \mathcal{A} = \mathbf{1}, \quad (\text{C3})$$

with  $\mathcal{A}^\dagger$  being another square root, then from Eq. (46) and Hermiticity of  $\mathcal{F}_q$  it will trivially follow that the matrix  $\tilde{\mathcal{F}}_q = \mathcal{A}^\dagger \mathcal{F}_q \mathcal{A}$  will be both Hermitian and symmetric. It is easy to check that formula (49) gives the square root of  $\Pi$  with all required properties.

(2) In the  $L = K - 1$  case  $\Pi$  is given by (50), and hence we can write (46) in the form

$$\begin{aligned} \tilde{\mathcal{F}}_q^{\alpha\alpha'} &= (\tilde{\Pi} \tilde{\mathcal{F}}_q^T \tilde{\Pi})^{\alpha\alpha'}, \quad \mathcal{F}_q^{K-1K-1} = \mathcal{F}_q^{K-1K-1}, \\ \mathcal{F}_q^{\alpha K-1} &= \sum_{\alpha'} f_q^{-1} \tilde{\Pi}^{\alpha\alpha'} \mathcal{F}_q^{K-1\alpha'}, \end{aligned} \quad (\text{C4})$$

where  $\alpha, \alpha' = 0, \dots, K - 2$ . Thus, for submatrices  $\tilde{\mathcal{F}}_q$  and  $\tilde{\Pi}$  of the size  $K - 2$  we get the problem analogous to that considered in (1). The matrix element  $\mathcal{F}_q^{K-1K-1}$  is explicitly real. Finally, for the remaining nondiagonal elements similarly one can check that the combination

$$\tilde{\mathcal{F}}_q^\alpha \equiv \sum_{\alpha'} f_q^{1/2} (\tilde{\mathcal{A}}^\dagger)^{\alpha\alpha'} \mathcal{F}_q^{\alpha' K-1} \quad (\text{C5})$$

is real due to (C4). Combining our results together we get the compact form of the unitary transformation (51).

- 
- [1] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1988), 3rd printing.
- [2] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford Science, New York, 1989).
- [3] D. J. Flory, *Statistical Mechanics of Chain Molecules* (Wiley, New York, 1979).
- [4] J. des Cloizeaux and G. Jannink, *Polymers in Solution* (Clarendon, Oxford, 1990).
- [5] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (American Institute of Physics, New York, 1994).
- [6] P. G. de Gennes, *J. Phys. (Paris) Lett.* **46**, L639 (1985).
- [7] A. Yu. Grosberg and D. V. Kuznetsov, *Macromolecules* **26**, 4249 (1993).
- [8] B. Ostrovsky and Y. Bar-Yam, *Comput. Polym. Sci.* **3**, 9 (1993); M. A. Smith, Y. Bar-Yam, Y. Rabin, B. Ostrovski, C. A. Bennett, N. Margolus, and T. Toffoli, *ibid.* **2**, 165 (1992).
- [9] E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson, *J. Chem. Phys.* **102**, 1816 (1995).
- [10] Yu. A. Kuznetsov, E. G. Timoshenko, and K. A. Dawson, *J. Chem. Phys.* **104**, 3338 (1996).
- [11] *Protein Folding*, edited by T.E. Creighton (W. H. Freeman, New York, 1992); *New Developments in Theoretical Studies of Proteins*, edited by R. Elber (World Scientific, Singapore, 1994).
- [12] C. Tanford, *The Hydrophobic Effect* (Wiley, New York, 1980); K. A. Dill, *Biochemistry* **29**, 7133 (1990); S. P. Obukhov, *J. Phys. A* **19**, 3655 (1986); C. D. Sfatos, A. M. Gutin, and E. I. Shakhnovich, *Phys. Rev. E* **48**, 465 (1993).
- [13] Yu. A. Kuznetsov, E. G. Timoshenko, and K. A. Dawson, *J. Chem. Phys.* **103**, 4807 (1995).
- [14] D. Bratko and K. A. Dawson, *J. Chem. Phys.* **99**, 5352 (1993).
- [15] E. G. Timoshenko and K. A. Dawson, *Phys. Rev. E* **51**, 492 (1995).
- [16] Actually, such counting gives the maximal number of independent combinations of the zero modes. Due to the inversion symmetry (see Sec. III) some of them may become linear combinations of others.
- [17] To generalize our method to the case of an open polymer with  $M \neq 1$  it is sufficient to replace Eq. (9) by the appropriate Fourier transform with  $f_n^{(q)} = \cos[\pi q(n + 1/2)/M]$ .
- However, for very long chains within this approximation the boundary effects are insignificant as we have shown in Ref. [10]. Moreover, the equations for an open chain become equivalent to those of a ring chain with twice as many units and appropriately redefined virial coefficients  $u_{\text{ring}}^{(L)} = 2^{1-L} u_{\text{open}}^{(L)}$ .
- [18] More precisely, the canonical rank  $L = K - 1$  is attributed to a sequence such that both  $L$  and  $K - L$  are odd numbers. In this case it is impossible to make the rank even by any shift of the chain reference point, and  $K$  is even. If both numbers  $L, K - L$  are even, the canonical rank is  $L = 0$ . If one of the numbers  $L, K - L$  is odd and another one is even, the canonical rank is not uniquely defined, as it could be made either  $L = 0$  or  $L = K - 1$  by transformations (39). This only happens for odd  $K$ . In such a case one may choose to work with either of the two possible canonical ranks  $L$ , and we call such sequences "odd." Hence every chain sequence could be one of the four possible types: irreducible (if the inversion symmetry does not exist, for example, *abaabb*), canonical rank zero (e.g., *abba*), canonical rank  $K = 1$  (e.g., *ab*) or "odd" (e.g., *aba*).
- [19] We note that this unitary transformation is not uniquely defined. To illustrate this we consider an alternative expression to Eq. (51). Anyway, the resulting self-consistency equations in terms of the matrix elements will be precisely the same for any such transformation.
- [20] Yu. A. Kuznetsov, E. G. Timoshenko, and K. A. Dawson, *J. Chem. Phys.* **104**, 336 (1996).
- [21] N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **103**, 4732 (1995).
- [22] C. D. Sfatos, A. M. Gutin, and E. I. Shakhnovich, *Phys. Rev. E* **50**, 2898 (1994); V. S. Pande, A. Yu. Grosberg, and T. Tanaka, *Phys. Rev. E* **51**, 3381 (1995); *J. Chem. Phys.* **103**, 9482 (1995).
- [23] T. Garel, L. Leibler, and H. Orland (unpublished).
- [24] This choice corresponds to  $\hat{u}_{11}^{(2)} \ll \hat{u}^{(3)}$ , which actually appears in the equations.
- [25] Yu. A. Kuznetsov, E. G. Timoshenko, and K. A. Dawson (unpublished).
- [26] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [27] E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson (unpublished).