Conformational properties of randomly flexible heteropolymers

Pallavi Debnath and Binny J. Cherayil

Department of Inorganic and Physical Chemistry, Indian Institute of Science, Bangalore-560012, India

Random copolymers made up of subunits with arbitrary degrees of flexibility are useful as models of biomolecules with different kinds of secondary structural motifs. We show that the mean square end-to-end distance $\langle R^2 \rangle$ of a two-letter $A-B$ random heteropolymer in which the constituent polymeric subunits are represented as continuum wormlike chains and the randomness is described by the two-state Markov process introduced by Fredrickson, Milner, and Leibler [Macromolecules **25**, 6341 (1992)] can be obtained in closed form. The expression for $\langle R^2 \rangle$ is a function of several parameters, including the number n of subunits, the fraction f of one kind of subunit, the persistence lengths l_A and l_B of the two subunits, and the degree of correlation λ between successive subunits. The variation of $\langle R^2 \rangle$ with each of these parameters is discussed.

I. INTRODUCTION

Many of the most important biological macromolecules—proteins and polynucleotides in particular—are copolymers of a small number of chemically distinctive monomeric subunits whose arrangement along the backbone of the polymer often leads to quite specific three-dimensional geometries. The relation between sequence and structure in such molecules is generally complex, but it can sometimes be inferred from the study of *minimal* models of the molecule. Random heteropolymers of just two elementary subunits A and B have proven to be especially useful as prototypes of the more complex heteropolymers typical of biological systems.[1] But there are certain recurring structural motifs in biomolecules that are incorporated into these minimal models only with difficulty. Helices are an example. In isolation, helices are frequently represented as wormlike chains using the Kratky–Porod model,[2,3] or one of its many variants.[4] When helices are part of a larger complex that may contain extended regions of complete flexibility distributed at random, however, an analogous description of the resulting *randomly* semiflexible polymer is less readily developed.

To varying levels of sophistication, there do exist calculations of the combined effects of stiffness and backbone disorder[5–7] (as well as of stretching forces, in some cases[8]) on the conformational properties of polymers. But an *exact* treatment (along the lines of Saitô, Takahashi, and Yunoki's path integral approach to the Kratky–Porod wormlike chain[3]), which could, potentially, exploit powerful field theoretic techniques to address questions that might otherwise prove intractable, does not appear to have been formulated. Such treatments, being exact, could provide valuable reference points for the development of approximate theories of related, more complicated systems when exact solutions are unavailable.

These considerations suggest exploring the utility of the STY methodology in analyzing the behavior of chains made up of randomly distributed semiflexible segments. Accordingly, in this paper, we apply the methodology to calculate the average radial dimensions of a polymer in which A and B "prepolymers" of arbitrary stiffness are arranged at random along the chain backbone. The sequence of A's and B's along the chain is assumed to be governed by the statistics of the same two-state Markov process that was used by Fredrickson, Milner, and Leibler (FML)[7] to analyze microphase ordering in random block copolymer melts. We show that with this choice of disorder, the STY model of the random heteropolymer so defined, also admits of an exact solution. Moreover, we find that the results are independent of whether the disorder is regarded as annealed or quenched.

The following section introduces the model, and sets up the expressions needed to calculate the mean square end-to-end distance of the chain. Section III uses the STY propagator to reduce these expressions to simpler analytical forms, which are then averaged over the sequence distribution, as discussed in Sec. IV. The final expressions obtained by this averaging operation are extremely lengthy, so all but the most pertinent results are relegated to the Appendix. Section V discusses these results in terms of the various parameters that define the model.

II. THE MODEL

We are interested in the conformational properties (specifically the mean square end-to-end distance) of a chain of length M made up of a sequence of n polymeric subunits ("prepolymers"). Each prepolymer is of length N (hence $nN=M$), and is of one of two kinds: A or B. The prepolymers are regarded as semiflexible, so in a continuum representation of the chain, they can be completely characterized by the set of unit tangent vectors $\mathbf{u}(\tau)$ at each of the points τ along the backbone. Within this representation, using units in which $k_B T = 1$, the Hamiltonian H is given by

$$H = \frac{1}{2}\epsilon_A \sum_{i=1}^{n} \delta_{\theta_{i},1} \int_{(i-1)N}^{iN} d\tau\, \dot{\mathbf{u}}^2(\tau)$$

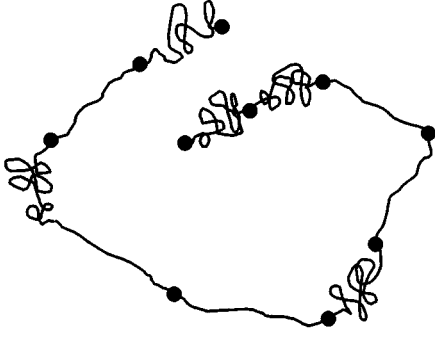$$+ \frac{1}{2}\epsilon_B \sum_{i=1}^{n} \delta_{\theta_{i},-1} \int_{(i-1)N}^{iN} d\tau\, \dot{u}^2(\tau), \tag{1}$$

FIG. 1. A sketch of one possible realization of an $A-B$ heteropolymer in which the A and B segments have different degrees of stiffness.

where θ_i is a discrete random variable that takes on the values ± 1, $+1$ when the ith prepolymer is of type A and -1 when it is of type B; ϵ_A and ϵ_B are the energies of bending of the segments A and B, respectively, and can be identified with the persistence lengths l_A and l_B of these segments; and δ is the Kronecker delta. Figure 1 is a sketch of one possible realization of the copolymer sequence when the A and B prepolymers have fairly different degrees of stiffness. In the sketch, the A prepolymers are "coil-like," while the B prepolymers are "helixlike." Had the A segments been completely rigid, the chain would have corresponded to the broken rod model studied by Muroga et al.[5]

A simple transformation[9] of the Kronecker deltas allows H to be rewritten more succinctly as

$$H = \frac{1}{2} \sum_{i=1}^{n} \Delta_i \int_{(i-1)N}^{iN} d\tau \dot{u}^2(\tau), \qquad (2)$$

where

$$\Delta_i = \tfrac{1}{2}(\epsilon_A + \epsilon_B) + \tfrac{1}{2}(\epsilon_A - \epsilon_B)\theta_i \qquad (3)$$

$$\equiv D_1 + D_2 \theta_i. \qquad (4)$$

The vectorial distance $\mathbf{R}$ from one end of the chain to the other is the sum of the end-to-end vectors $\mathbf{R}_i$ of each of the prepolymers, i.e.,

$$\mathbf{R} = \sum_{i=1}^{n} \mathbf{R}_i, \qquad (5)$$

where $\mathbf{R}_i$ itself is given by

$$\mathbf{R}_i = \int_{(i-1)N}^{iN} d\tau \mathbf{u}(\tau). \qquad (6)$$

Thus, the mean square end-to-end distance $\langle R^2 \rangle$ can be written in the form

$$\langle R^2 \rangle = S_1 + 2(S_2 + S_3), \qquad (7)$$

where

$$S_1 = \sum_{i=1}^{n} \langle R_i^2 \rangle$$

$$= \sum_{i=1}^{n} \int_{(i-1)N}^{iN} d\tau_1 \int_{(i-1)N}^{iN} d\tau_2 \langle \mathbf{u}(\tau_1) \cdot \mathbf{u}(\tau_2) \rangle, \qquad (8)$$

$$S_2 = \sum_{i=1}^{n-1} \langle \mathbf{R}_i \cdot \mathbf{R}_{i+1} \rangle$$

$$= \sum_{i=1}^{n-1} \int_{(i-1)N}^{iN} d\tau_1 \int_{iN}^{(i+1)N} d\tau_2 \langle \mathbf{u}(\tau_1) \cdot \mathbf{u}(\tau_2) \rangle, \qquad (9)$$

$$S_3 = \sum_{i=1}^{n-2} \sum_{j=i+2}^{n} \langle \mathbf{R}_i \cdot \mathbf{R}_j \rangle$$

$$= \sum_{i=1}^{n-2} \sum_{j=i+2}^{n} \int_{(i-1)N}^{iN} d\tau_1 \int_{(j-1)N}^{jN} d\tau_2 \langle \mathbf{u}(\tau_1) \cdot \mathbf{u}(\tau_2) \rangle. \qquad (10)$$

The angular brackets in Eqs. (7)–(10) denote an average both over the conformational degrees of freedom of the chain as well as the distribution of the discrete random variables θ_i. From these expressions the calculation of $\langle R^2 \rangle$ is seen to reduce essentially to the calculation of the "bond" correlation function $\langle \mathbf{u}(\tau_1) \cdot \mathbf{u}(\tau_2) \rangle$.

III. EVALUATION OF THE BOND CORRELATION FUNCTION

In general, the correlation function of the tangent vectors can be written in the form

$$\langle \mathbf{u}(\tau_1) \cdot \mathbf{u}(\tau_2) \rangle = \frac{1}{Q} \int \mathcal{D}[\mathbf{u}(\tau)] \mathbf{u}(\tau_1) \cdot \mathbf{u}(\tau_2) e^{-H}, \qquad (11)$$

where H is given by Eq. (1), Q is the partition function, defined as

$$Q = \int \mathcal{D}[\mathbf{u}(\tau)] e^{-H}, \qquad (12)$$

and $\mathcal{D}[\mathbf{u}(\tau)]$ is the functional integral measure on the space of functions $\mathbf{u}(\tau)$. From the general approach to the evaluation of functional integrals described, for example, in Ref. 10, Eq. (11) can be reduced to a product of ordinary integrals involving a Green's function $G(\mathbf{u}, \mathbf{u}'|\tau - \tau')$, which describes the probability density that a monomer at the point τ' on the chain has the bond orientation $\mathbf{u}'$ if the monomer at τ has the orientation $\mathbf{u}$. When these points are located on the jth and ith prepolymers, respectively, the expression for the bond correlation function can be shown to be given by

$$\langle \mathbf{u}(\tau_1)\cdot\mathbf{u}(\tau_2)\rangle$$

$$= \frac{1}{Q}\int d\mathbf{u}_0\int d\mathbf{u}_1\cdots\int d\mathbf{u}_{i-1}\int d\mathbf{u}\int d\mathbf{u}_i\cdots\int d\mathbf{u}_{j-1}$$

$$\times\int d\mathbf{u}'\int d\mathbf{u}_j\cdots\int d\mathbf{u}_{n-1}$$

$$\times\int d\mathbf{u}_n\mathbf{u}\cdot\mathbf{u}'\,G(\mathbf{u}_0,\mathbf{u}_1|N)$$

$$\times G(\mathbf{u}_1,\mathbf{u}_2|N)\cdots G(\mathbf{u}_{i-1},\mathbf{u}|\tau_1-(i-1)N)$$

$$\times G(\mathbf{u},\mathbf{u}_i|iN-\tau_1)\cdots G(\mathbf{u}_{j-1},\mathbf{u}'|\tau_2$$

$$-(j-1)N)G(\mathbf{u}',\mathbf{u}_j|jN-\tau_2)\cdots G(\mathbf{u}_{n-1},\mathbf{u}_n|N), \quad (13)$$

where, in general,

$$G(\mathbf{u}_{k-1},\mathbf{u}_k|\tau_2-\tau_1)=\int_{\mathbf{u}(\tau_1)=\mathbf{u}_{k-1}}^{\mathbf{u}(\tau_2)=\mathbf{u}_k}\mathcal{D}[\mathbf{u}(\tau)]$$

$$\times\exp\left[-\frac{1}{2}\Delta_k\int_{\tau_1}^{\tau_2}d\tau\dot{\mathbf{u}}^2(\tau)\right]. \quad (14)$$

In terms of this distribution function, the partition function Q can be similarly written as

$$Q=\int d\mathbf{u}_0\prod_{i=1}^{n}\int d\mathbf{u}_i G(\mathbf{u}_{i-1},\mathbf{u}|iN-(i-1)N). \quad (15)$$

Adopting the STY model of the semiflexible chain[3] (which ensures inextensibility of the chain through the constraint $|\mathbf{u}(\tau)|=1$), one can determine the Green's function as an expansion in spherical harmonics:

$$G(\mathbf{u}_{k-1},\mathbf{u}_k|\tau_2-\tau_1)=\sum_{m,n}e^{-a_n|\tau_2-\tau_1|/2\Delta_k}$$

$$\times Y_{m,n}^{*}(\mathbf{u}_{k-1})Y_{m,n}(\mathbf{u}_k), \quad (16)$$

where $a_n\equiv n(n+1)$. These spherical harmonics satisfy the relations

$$\int d\mathbf{u}Y_{m_1,n_1}^{*}(\mathbf{u})Y_{m_2,n_2}(\mathbf{u})=\delta_{m_1,m_2}\delta_{n_1,n_2}, \quad (17)$$

$$\int d\mathbf{u}Y_{m,n}(\mathbf{u})=\sqrt{4\pi}\delta_{m,0}\delta_{n,0}, \quad (18)$$

$$Y_{0,0}=\frac{1}{\sqrt{4\pi}}, \quad (19)$$

from which it immediately follows that $Q=4\pi$, independent of the random variable θ_i. This implies that there is no distinction here between quenched and annealed disorder.

To evaluate the functions S_1, S_2, and S_3 that are needed in determining $\langle R^2\rangle$, it is helpful to refer to Fig. 2(a), which is a diagrammatic representation of the bond correlation function $\langle\mathbf{u}(\tau_1)\cdot\mathbf{u}(\tau_2)\rangle$. The straight line segments in this figure stand for the Green's function [Eq. (14)] for the section of prepolymer of type A or B that lies between the junction points at the indicated contour positions. The crosses mark the locations of the points τ_1 and τ_2 that appear in Eq. (11), defining the correlation function; these points may lie
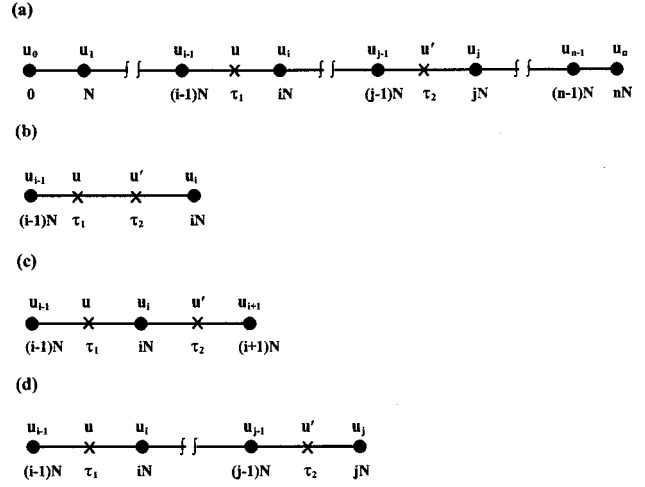

FIG. 2. (a) A schematic representation of an $A-B$ heteropolymer of n prepolymer segments showing coordinates relevant to the calculation of the end-to-end distance. Each straight line segment terminated by full circles is an A or a B prepolymer of contour length N. Letters above the circles and crosses stand for the unit tangent vectors at the contour positions indicated by the letters below these symbols. Squiggles denote sections of the chain that are not shown in the diagram. (b) The diagrammatic representation of the integral D_1 [Eq. (20)]. (c) The diagrammatic representation of the integral D_2 [Eq. (21)]. (d) The diagrammatic representation of the integral D_3 [Eq. (22)].

anywhere on the backbone of the chain. Squiggles indicate prepolymer segments that are not shown. Integrations are understood to be carried out over the vectorial positions of all junction points (including those labeled by τ_1 and τ_2).

When the spherical harmonic expansion of the Green's function is substituted into Eq. (13), and the integrations carried out using Eqs. (17)–(19), $\langle\mathbf{u}(\tau_1)\cdot\mathbf{u}(\tau_2)\rangle$ in S_1 is seen to simplify to the diagram shown in Fig. 2(b). If the algebraic expression for this diagram is denoted D_1, one can show that

$$D_1=4\pi\int d\mathbf{u}_1\int d\mathbf{u}_2\mathbf{u}_1\cdot\mathbf{u}_2$$

$$\times\sum_{n,m}\exp(-a_n|\tau_2-\tau_1|/2\Delta_i)Y_{n,m}^{*}(\mathbf{u}_1)Y_{n,m}(\mathbf{u}_2). \quad (20)$$

Similarly, $\langle\mathbf{u}(\tau_1)\cdot\mathbf{u}(\tau_2)\rangle$ in S_2, reduces to D_2 [Fig. 2(c)], which is given by

$$D_2=4\pi\int d\mathbf{u}_1\int d\mathbf{u}_2\mathbf{u}_1\cdot\mathbf{u}_2\sum_{n,m}\exp(-a_n|iN-\tau_1|/2\Delta_i)$$

$$\times\exp(-a_n|\tau_2-iN|/2\Delta_{i+1})Y_{n,m}^{*}(\mathbf{u}_1)Y_{n,m}(\mathbf{u}_2). \quad (21)$$

Likewise, but after somewhat more algebra, $\langle\mathbf{u}(\tau_1)\cdot\mathbf{u}(\tau_2)\rangle$ in S_3, reduces to D_3 [Fig. 2(d)], which is given by

$$D_3 = 4\pi \int d\mathbf{u}_1 \int d\mathbf{u}_2 \mathbf{u}_1 \cdot \mathbf{u}_2 \sum_{n,m} \exp(-a_n |iN - \tau_1|/2\Delta_i)$$

$$\times \exp\left(-a_n N \sum_{l=1}^{j-i-1} \frac{1}{2\Delta_{i+l}}\right)$$

$$\times \exp(-a_n |\tau_2 - (j-1)N|/2\Delta_j) Y_{n,m}^*(\mathbf{u}_1) Y_{n,m}(\mathbf{u}_2). \tag{22}$$

This expression applies specifically to the case where the jth prepolymer satisfies $j \geq i + 2$.

The calculation of the average over the conformational degrees of freedom is finally completed by evaluating the integrals over $\mathbf{u}_1$ and $\mathbf{u}_2$ in Eqs. (20)–(22). Noting that $\mathbf{u}_1 \cdot \mathbf{u}_2 = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2)$, one can use standard results from the theory of spherical harmonics to show that[11]

$$\int d\mathbf{u}_1 \int d\mathbf{u}_2 \mathbf{u}_1 \cdot \mathbf{u}_2 Y_{n,m}^*(\mathbf{u}_1) Y_{n,m}(\mathbf{u}_2)$$

$$= \frac{4\pi}{3} \delta_{n,1}(\delta_{m,-1} + \delta_{m,0} + \delta_{m,1}). \tag{23}$$

When the above equation is used in Eqs. (20)–(22), and the resulting expressions then substituted into Eq. (7), we obtain

$$\langle R^2 \rangle = 2N \sum_{i=1}^{n} \Delta_i \left[1 - \frac{\Delta_i}{N}(1 - e^{-N/\Delta_i}) \right]$$

$$+ 2 \sum_{i=1}^{n-1} \Delta_i \Delta_{i+1}(1 - e^{-N/\Delta_i})(1 - e^{-N/\Delta_{i+1}})$$

$$+ 2 \sum_{i=1}^{n-2} \sum_{j=i+2}^{n} \Delta_i \Delta_j \exp\left[-N \sum_{l=1}^{j-i-1} \frac{1}{\Delta_{i+1}}\right]$$

$$\times (1 - e^{-N/\Delta_i})(1 - e^{-N/\Delta_j}). \tag{24}$$

When this expression is averaged over the sequence distribution (as discussed in the following section), the desired disorder-averaged end-to-end distance is obtained.

IV. AVERAGE OVER THE SEQUENCE DISTRIBUTION

In the absence of sequence disorder, there is no distinction between A and B segments, so the model describes a homopolymer of n semiflexible prepolymer segments whose stiffness can be characterized by a single persistence length l, where $l = \epsilon_A = \epsilon_B$. It is easily verified that in this limit, Eq. (24) for $\langle R^2 \rangle$ correctly reproduces the Kratky–Porod description of the chain. In particular, when $l \gg 1$, $\langle R^2 \rangle$ scales as $n^2 N^2$, whereas when $l \ll 1$, $\langle R^2 \rangle$ scales as Ml^2.

When the chain is a random heteropolymer, an average over the sequence distribution must be carried out explicitly to produce the final expression for $\langle R^2 \rangle$. To perform this average, we adopt the model introduced by Fredrickson, Milner, and Leibler[7] to discuss phase separation in random copolymer blends. In this model, the probability that ith prepolymer in the chain is of a given type is assumed to be determined solely by the chemical identity of the immediately preceding prepolymer and no others. Thus, the probability of realizing a given sequence of A's and B's is deter-

mined, in general, by a set of four conditional probabilities p_{KL}, K, $L = A$, B, where p_{KL} is conditional probability of observing K given L. If it is further assumed that this sequence is the end result of a living polymerization process under steady state conditions, the p_{KL} can be expressed in terms of the mole fractions f and $1 - f$ that define the composition of A and B in the initial reaction mixture (and in the chain that is generated thereby.) These assumptions, along with the Markov condition, establish that only *two* parameters need be specified to fix the overall average sequence distribution: one is f itself, and the other is the nontrivial eigenvalue λ of the matrix of conditional probabilities. In terms of these parameters, the following relations for the p_{KL} can be derived:[7]

$$p_{AA} = f(1 - \lambda) + \lambda, \tag{25}$$

$$p_{BB} = f(\lambda - 1) + 1, \tag{26}$$

$$p_{AB} = 1 - p_{AA}, \tag{27}$$

$$p_{BA} = 1 - p_{BB}. \tag{28}$$

Physically, the parameter λ is a measure of the extent of "blockiness" (to use FML's phrase) of the chain; it can be shown to assume values between -1 and $+1$. The limit $\lambda \to -1$ describes a chain in which A and B prepolymers tend to succeed each other in alternation, the limit $\lambda \to +1$ describes a chain in which A's tend to succeed A's, and B's tend to succeed B's, and the limit $\lambda = 0$ describes a chain in which the A's and B's follow each other entirely randomly.

To return to Eq. (24), we see that the sequence average of the end-to-end distance requires averages over quantities like $\exp(-N/\Delta_i)$. To perform such averages, recall that $\Delta_i = D_1 + D_2 \theta_i$, where D_1 and D_2 are defined in Eq. (4) in terms of the bending energies ϵ_A and ϵ_B. It follows therefore that

$$\frac{1}{\Delta_i} = \frac{D_1}{D_1^2 - D_2^2} - \frac{D_2}{D_1^2 - D_2^2} \theta_i. \tag{29}$$

Defining $\alpha = D_1/(D_1^2 - D_2^2)$ and $\beta = D_2/(D_1^2 - D_2^2)$, we now have

$$\langle e^{-N/\Delta_i} \rangle = e^{-N\alpha} \langle e^{-N\beta\theta_i} \rangle, \tag{30}$$

where the angular brackets now refer to the average over the distribution of the θ_i's. Since θ_i takes the values ± 1, it is easy to show that

$$\langle e^{-N\beta\theta_i} \rangle = \cosh N\beta + \langle \theta_i \rangle \sinh N\beta. \tag{31}$$

The average $\langle \theta_i \rangle$ is obtained from the relation $\langle \theta_i \rangle = \Sigma_{\theta = \pm 1} \theta p_s(\theta)$, where $p_s(+1) = f$ and $p_s(-1) = 1 - f$. Clearly, $\langle \theta_i \rangle = 2f - 1$. From this result it is easy to show that

$$\langle S_1 \rangle = 2nN\left[D_1 + (2f-1)D_2 - \frac{D_1^2 + D_2^2}{N} \right.$$

$$\times \{1 - e^{-N\alpha}(\cosh N\beta + (2f-1)\sinh N\beta)\}$$

$$- \frac{2D_1 D_2}{N}\{2f - 1 - e^{-N\alpha}(\sinh N\beta$$

$$\left. + (2f-1)\cosh N\beta)\} \right]. \tag{32}$$

The sequence averaged value of S_2 is similarly calculated as

$$\langle S_2 \rangle = 2(n-1)D_1^2(1-e^{-N\alpha}\cosh N\beta)^2$$

$$\times \left[\left\{ 1 - \frac{AD_2}{D_1} - (2f-1)\left(A - \frac{D_2}{D_1}\right) \right\}^2 \right.$$

$$\left. + 4f(1-f)\lambda\left(A - \frac{D_2}{D_1}\right)^2 \right], \tag{33}$$

where

$$A \equiv \frac{e^{-N\alpha}\sinh N\beta}{1-e^{-N\alpha}\cosh N\beta}. \tag{34}$$

The calculation of $\langle S_3 \rangle$ is much less trivial, but it can be done analytically. Details of the calculation are provided in Appendix A. Here we quote the result in terms of four other averages $\langle S_{31} \rangle$, $\langle S_{32} \rangle$, $\langle S_{33} \rangle$, and $\langle S_{34} \rangle$, whose complete expressions in terms of the various parameters of the model are given in Eqs. (A21), (A24), (A27), and (A30):

$$\langle S_3 \rangle = 2\sum_{i=1}^{n-2}\sum_{j=i+2}^{n} D_1^2(2-e^{-N\alpha}\cosh N\beta)^2$$

$$\times \left[\left(1 - \frac{AD_2}{D_1}\right)^2 e^{-(j-i-1)N\alpha}\langle S_{31} \rangle - \left(1 - \frac{AD_2}{D_1}\right) \right.$$

$$\times \left(A - \frac{D_2}{D_1}\right)e^{-(j-i-1)N\alpha}(\langle S_{32} \rangle + \langle S_{33} \rangle)$$

$$\left. + \left(A - \frac{D_2}{D_1}\right)^2 e^{-(j-i-1)N\alpha}\langle S_{34} \rangle \right]. \tag{35}$$

After substituting the equations for $\langle S_{31} \rangle$, $\langle S_{32} \rangle$, $\langle S_{33} \rangle$, and $\langle S_{34} \rangle$ into the above equation, the sums become trivial, and can be done at once.

V. DISCUSSION

Equation (7), along with Eqs. (32) and (33), and Eq. (35) (after carrying out the summations), is the desired expression for the disorder averaged mean square end-to-end distance of the given $A-B$ copolymer, but it is far too lengthy and complicated to be particularly perspicuous on its own. It is therefore depicted graphically in a series of figures (3–6) that highlight its behavior in terms of one or other of the parameters that $\langle R^2 \rangle$ depends on. These parameters are the length N of the prepolymer, the number n of prepolymers in the chain as a whole, the fraction f of A-type prepolymers, the extent of blockiness λ, and the persistence lengths l_A and l_B of the A and B prepolymers, respectively. As a matter of convenience, N is kept constant throughout. Although f and λ are regarded as free parameters, the fact that the conditional probabilities in Eqs. (25)–(28) are constrained to lie between 0 and 1 implies that f and λ can only be chosen within certain limits; these limits are always respected when numerical values are assigned to the parameters.

We plot $\alpha_R \equiv \langle R^2 \rangle / M^2$ against the dimensionless inverse persistence length $1/l^* \equiv N/\epsilon_B \equiv N/l_B$ for four different values of one other parameter, all other parameters staying the same. A fifth curve is included for reference: this is the varia-
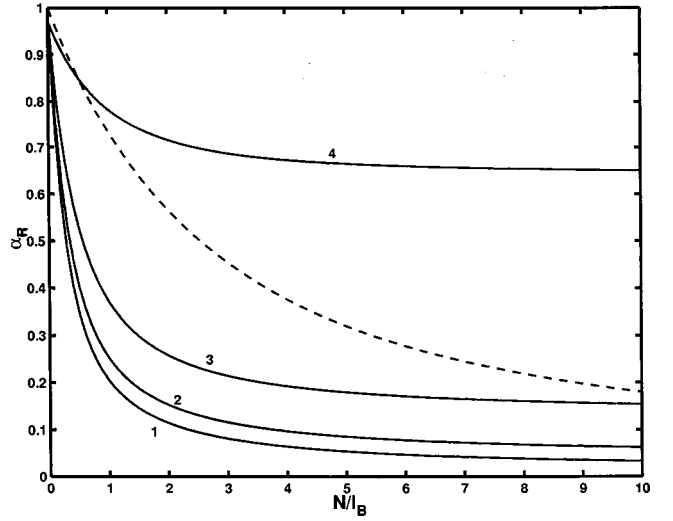

FIG. 3. Full curves are the variation of the parameter $\alpha_R \equiv \langle R^2 \rangle / M^2$ with the dimensionless inverse persistence length N/l_B as a function of the fraction f of A at fixed values of λ, N/l_A and n (0.0, 0.01, and 10, respectively.) The dashed curve is the Kratky–Porod expression for α_R of a homopolymer of length M derived from Eq. (36). Curves 1, 2, 3, and 4 correspond, respectively, to $f=0.1$, 0.25, 0.5, and 0.9.

tion of α_R as a function of M/l for a semiflexible *homopolymer* of contour length M as calculated with the following Kratky–Porod expression:[3]

$$\langle R^2 \rangle = Ml\left[1 - \frac{1}{2M/l}(1-e^{-2M/l})\right]. \tag{36}$$

This expression (which is reproduced by our model in the homopolymer limit) yields the results $\alpha_R \to 1$ as $l \gg 1$ and $\alpha_R \to 0$ as $l \ll 1$.

Figure 3 shows the variation of α_R [as determined from Eqs. (7) along with (32), (33) and (35)] with $1/l^*$ (the full lines) for four different values of f (0.10, 0.25, 0.50, and 0.90) at the following values of the other parameters: λ
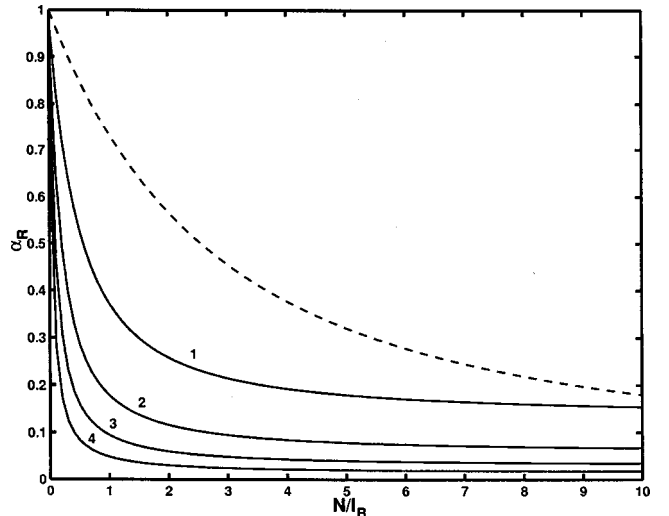

FIG. 4. Variation of α_R with N/l_B as a function of n at fixed values of f (0.5), λ(0.0), and N/l_A (0.01). Curves 1, 2, 3, and 4 correspond to $n=10$, 25, 50, and 100, respectively. The dashed curve is the same Kratky–Porod result shown in Fig. 3.
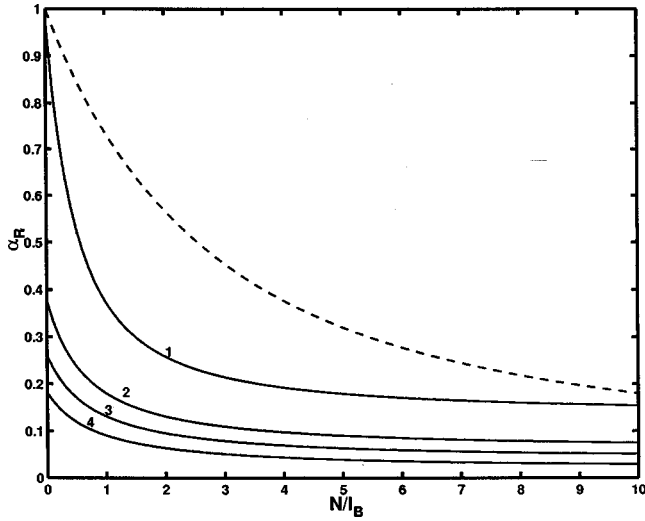
FIG. 5. Variation of α_R with N/l_B as a function of N/l_A at fixed values of f (0.5), λ (0.0), and n (10). Curves 1, 2, 3, and 4 correspond to $N/l_A = 0.01$, 1.0, 2.0, and 5.0, respectively. The dashed curve is the same Kratky–Porod result shown in Fig. 3.

$=0.0$, $N/l_A = 0.01$ and $n = 10$. The dashed line is the Kratky–Porod result derived from Eq. (36). Flexibility increases from left to right along the abscissa. In general, N/l_A values less than unity correspond to prepolymer segments that are semiflexible or rigid, so the choice $N/l_A = 0.01$ indicates that A is relatively stiff. The choice $\lambda = 0.0$ indicates that chain is an ideal random copolymer. The figure shows that principal effect of increasing the proportion of A in the chain is to render it increasingly *inflexible*, as one would expect. At the smallest fractions of A, however, there is a fairly narrow range of N/l_B values over which increasing the stiffness of B leads to a fairly sharp rise in the stiffness of the chain as a whole. This is a trend that is repeated, to greater or less degree, in all the other figures.
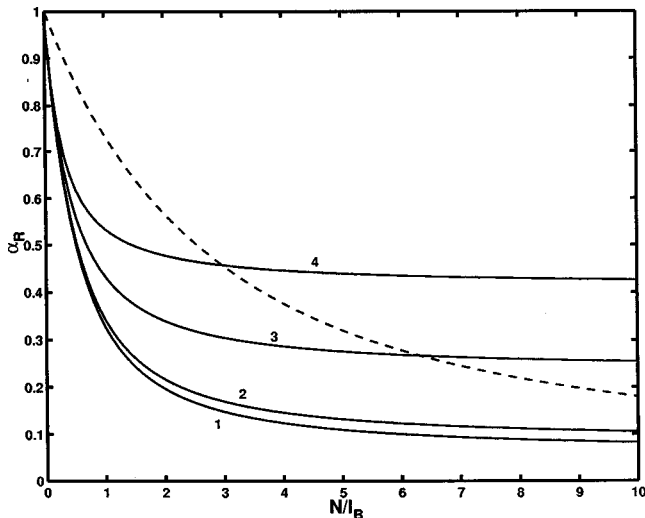

FIG. 6. Variation of α_R with N/l_B as a function of λ at fixed values of f (0.5), N/l_A (0.01), and n (10). Curves 1, 2, 3, and 4 correspond to $\lambda = -0.9$, -0.5, 0.5, and 0.9, respectively. The dashed curve is the same Kratky–Porod result shown in Fig. 3.

The sharpness of the change from flexible to rigid geometries is especially pronounced in Fig. 4, which shows the variation of α_R with $1/l^*$ for 4 different values of n (10, 25, 50, and 100) at the following fixed values of the other parameters: $f = 0.5$, $\lambda = 0.0$ and $N/l_A = 0.01$. Presumably, as $n \rightarrow \infty$, there is something akin to a genuine discontinuity between the flexible and rodlike configurations of the random heteropolymer.

The transition can be completely suppressed for certain ranges of parameter values, as illustrated in Fig. 5, which shows the variation of α_R with $1/l^*$ for 4 different values of N/l_A (0.01, 1, 2, and 5) at the following fixed values of the other parameters: $f = 0.5$, $\lambda = 0.0$ and $n = 10$. At the largest values of N/l_A (1, 2 and 5), corresponding to the greatest degree of conformational flexibility, the chain never attains more than about 40% of its full extension.

The blockiness parameter λ provides one final measure of conformational control, as illustrated in Fig. 6, which shows the variation of α_R with $1/l^*$ for four different values of this parameter (-0.9, -0.5, 0.5, and 0.9) at the following fixed values of the other parameters: $f = 0.5$, $N/l_A = 0.01$ and $n = 10$. The negative values of λ correspond to chains with the tendency to alternate between A and B segments, and such chains show the greatest degree of flexibility at N/l_B values greater than about 2. Below that number, when both A and B segments are relatively stiff, the conformation of the chain is increasingly insensitive to how frequently the two segments alternate with each other.

In conclusion, we have found an exact solution for the size of a random heteropolymer based on the STY model that highlights the interplay between flexibility and disorder in chain statistics. The analytical expression for $\langle R^2 \rangle$ that is derived from the model can be used to make rough estimates of the range of allowed sizes of chains of unknown structure but definite sequence.

APPENDIX A: CALCULATION OF $\langle S_3 \rangle$

The functions S_{31}, S_{32}, S_{33} and S_{34} that appear in S_3 [Eq. (35)] are defined as

$$S_{31} = \exp\left(N\beta \sum_{l=1}^{j-i-1} \theta_{i+l} \right), \tag{A1}$$

$$S_{32} = \theta_i \exp\left(N\beta \sum_{l=1}^{j-i-1} \theta_{i+l} \right), \tag{A2}$$

$$S_{33} = \exp\left(N\beta \sum_{l=1}^{j-i-1} \theta_{i+l} \right) \theta_j, \tag{A3}$$

$$S_{34} = \theta_i \exp\left(N\beta \sum_{l=1}^{j-i-1} \theta_{i+l} \right) \theta_j. \tag{A4}$$

By definition,

$$\langle S_{31} \rangle = \sum_{\theta_{i+1}} \sum_{\theta_{i+2}} \cdots \sum_{\theta_{j-1}} p_s(\theta_{i+1}) p(\theta_{i+2}|\theta_{i+1})$$
$$\times p(\theta_{i+3}|\theta_{i+2}) \cdots p(\theta_{j-1}|\theta_{j-2}) \exp(\mu\theta_{i+1})$$
$$\times \exp(\mu\theta_{i+2}) \cdots \exp(\mu\theta_{j-1}), \tag{A5}$$

where $p_s(\theta_i)$ is the equilibrium (i.e., steady-state) probability for the occurrence of the state θ_i, $p(\theta_{i+1}|\theta_i)$ is the conditional probability of seeing θ_{i+1} given θ_i, and μ is $N\beta$. One may verify that Eq. (A5) may be expressed in matrix form as

$$\langle S_{31}\rangle=(e^\mu \quad e^{-\mu})\underbrace{\begin{pmatrix}p_{AA} & p_{AB}\\ p_{BA} & p_{BB}\end{pmatrix}\begin{pmatrix}e^\mu & 0\\ 0 & e^{-\mu}\end{pmatrix}\cdots\begin{pmatrix}p_{AA} & p_{AB}\\ p_{BA} & p_{BB}\end{pmatrix}\begin{pmatrix}e^\mu & 0\\ 0 & e^{-\mu}\end{pmatrix}}_{j-i-2 \text{ factors}}\begin{pmatrix}p_A\\ p_B\end{pmatrix}, \tag{A6}$$

where $p_{AA}\equiv p(1|1)$, $p_{AB}\equiv p(1|-1)$, $p_{BA}\equiv p(-1|1)$ and $p_{AA}\equiv p(-1|-1)$. Defining the matrix

$$\mathbf{P}=\begin{pmatrix}p_{AA}e^\mu & p_{AB}e^{-\mu}\\ p_{BA}e^\mu & p_{BB}e^{-\mu}\end{pmatrix} \tag{A7}$$

it is seen that

$$\langle S_{31}\rangle=(e^\mu \quad e^{-\mu})\mathbf{P}^{j-i-2}\begin{pmatrix}p_A\\ p_B\end{pmatrix}. \tag{A8}$$

The eigenvalues Λ_1 and Λ_2 of $\mathbf{P}$ are given by

$$\Lambda_{1,2}=\tfrac{1}{2}(p_{AA}e^\mu+p_{BB}e^{-\mu})T_{1,2}, \tag{A9}$$

where

$$T_{1,2}=1\pm\sqrt{1-Q}, \tag{A10}$$

with

$$Q=\frac{4\lambda}{(p_{AA}e^\mu+p_{BB}e^{-\mu})^2}. \tag{A11}$$

The upper symbol in (A10) corresponds to T_1 while the lower symbol corresponds to T_2. Introduce a set of left eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$, and a set of right eigenvectors $\mathbf{y}_1$ and $\mathbf{y}_2$ satisfying

$$\mathbf{P}\mathbf{x}_1=\Lambda_1\mathbf{x}_1, \quad \mathbf{P}\mathbf{x}_2=\Lambda_2\mathbf{x}_2 \tag{A12}$$

and

$$\mathbf{y}_1^T\mathbf{P}=\mathbf{y}_1^T\Lambda_1, \quad \mathbf{y}_2^T\mathbf{P}=\mathbf{y}_2^T\Lambda_2. \tag{A13}$$

The right eigenvectors may be chosen as

$$\mathbf{x}_{1,2}=\begin{pmatrix}1\\ -(p_{AA}e^\mu-\Lambda_{1,2})/p_{AB}e^{-\mu}\end{pmatrix}. \tag{A14}$$

Similarly, the left eigenvectors may be chosen as

$$\mathbf{y}_{1,2}^T=(1, \quad -(p_{AA}e^\mu-\Lambda_{1,2})/p_{BA}e^\mu). \tag{A15}$$

For $i\neq j$, one may see from Eqs. (25)–(28) that $\mathbf{y}_i^T\mathbf{x}_j=0$. When $i=j$, we have

$$\mathbf{y}_i^T\mathbf{x}_i=1+\frac{1}{p_{AB}p_{BA}}(p_{AA}e^\mu-\Lambda_i)^2\equiv C_i, \quad i=1,2. \tag{A16}$$

Furthermore, if we define the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ according to

$$\mathbf{A}_i=\begin{pmatrix}1 & -\phi^*\\ -\phi & \phi\phi^*\end{pmatrix}, \quad i=1,2 \tag{A17}$$

where

$$\phi_i^*=(p_{AA}e^\mu-\Lambda_i)/p_{BA}e^\mu, \quad i=1,2 \tag{A18}$$

and

$$\phi_i=(p_{AA}e^\mu-\Lambda_i)/p_{AB}e^{-\mu}, \quad i=1,2, \tag{A19}$$

then, in general, for some integer m,[12]

$$\mathbf{P}^m=\frac{1}{C_1}\mathbf{A}_1\Lambda_1^m+\frac{1}{C_2}\mathbf{A}_2\Lambda_2^m. \tag{A20}$$

Hence,

$$\langle S_{31}\rangle=(e^\mu \quad e^{-\mu})\left[\frac{1}{C_1}\mathbf{A}_1\Lambda_1^{j-i-2}+\frac{1}{C_2}\mathbf{A}_2\Lambda_2^{j-i-2}\right]\begin{pmatrix}p_A\\ p_B\end{pmatrix}$$

$$=\frac{\Lambda_1^{j-i-2}}{C_1}[f-(1-f)\phi_1^*](e^\mu-\phi_1 e^{-\mu})$$

$$+\frac{\Lambda_2^{j-i-2}}{C_2}[f-(1-f)\phi_2^*](e^\mu-\phi_2 e^{-\mu}). \tag{A21}$$

In the same way

$$\langle S_{32}\rangle=\sum_{\theta_i}\sum_{\theta_{i+1}}\cdots\sum_{\theta_{j-1}}p_s(\theta_i)p(\theta_{i+1}|\theta_i)p(\theta_{i+2}|\theta_{i+1})$$

$$\cdots p(\theta_{j-1}|\theta_{j-2})\theta_i\exp(\mu\theta_{i+1})$$

$$\times\exp(\mu\theta_{i+2})\cdots\exp(\mu\theta_{j-1}), \tag{A22}$$

which in matrix notation is given by

$$\langle S_{32}\rangle=(e^\mu \quad e^{-\mu})\underbrace{\begin{pmatrix}p_{AA} & p_{AB}\\ p_{BA} & p_{BB}\end{pmatrix}\begin{pmatrix}e^\mu & 0\\ 0 & e^{-\mu}\end{pmatrix}\cdots\begin{pmatrix}p_{AA} & p_{AB}\\ p_{BA} & p_{BB}\end{pmatrix}\begin{pmatrix}e^\mu & 0\\ 0 & e^{-\mu}\end{pmatrix}\begin{pmatrix}p_{AA} & p_{AB}\\ p_{BA} & p_{BB}\end{pmatrix}}_{j-i-2 \text{ factors}}\begin{pmatrix}p_A\\ -p_B\end{pmatrix}, \tag{A23}$$

so using the result of Eq. (A20)

$$\langle S_{32}\rangle = (e^{\mu} \quad e^{-\mu})\left[\frac{1}{C_1}\mathbf{A}_1\Lambda_1^{j-i-2}+\frac{1}{C_2}\mathbf{A}_2\Lambda_2^{j-i-2}\right]$$

$$\times\begin{pmatrix}p_{AA}p_A-p_{AB}p_B\\p_{BA}p_A-p_{BB}p_B\end{pmatrix}$$

$$=\frac{\Lambda_1^{j-i-2}}{C_1}(e^{\mu}-\phi_1 e^{-\mu})$$

$$\times[fp_{AA}-(1-f)p_{AB}-\phi_1^*(fp_{BA}-(1-f)p_{BB})]$$

$$+\frac{\Lambda_2^{j-i-2}}{C_2}(e^{\mu}-\phi_2 e^{-\mu})[fp_{AA}-(1-f)p_{AB}$$

$$-\phi_2^*(fp_{BA}-(1-f)p_{BB})]. \tag{A24}$$

Similarly,

$$\langle S_{33}\rangle = \sum_{\theta_{i+1}}\sum_{\theta_{i+2}}\cdots\sum_{\theta_j}p_s(\theta_{i+1})p(\theta_{i+2}|\theta_{i+1})$$

$$\times p(\theta_{i+3}|\theta_{i+2})\cdots p(\theta_j|\theta_{j-1})\exp(\mu\theta_{i+1})$$

$$\times\exp(\mu\theta_{i+2})\cdots\exp(\mu\theta_{j-1})\theta_j, \tag{A25}$$

which in matrix notation is given by

$$\langle S_{33}\rangle = (1 \quad -1)\underbrace{\begin{pmatrix}p_{AA}&p_{AB}\\p_{BA}&p_{BB}\end{pmatrix}\begin{pmatrix}e^{\mu}&0\\0&e^{-\mu}\end{pmatrix}\cdots\begin{pmatrix}p_{AA}&p_{AB}\\p_{BA}&p_{BB}\end{pmatrix}\begin{pmatrix}e^{\mu}&0\\0&e^{-\mu}\end{pmatrix}}_{j-i-1\ \text{factors}}\begin{pmatrix}p_A\\p_B\end{pmatrix} \tag{A26}$$

$$=(1 \quad -1)\left[\frac{1}{C_1}\mathbf{A}_1\Lambda_1^{j-i-1}+\frac{1}{C_2}\mathbf{A}_2\Lambda_2^{j-i-1}\right]\begin{pmatrix}p_A\\p_B\end{pmatrix}$$

$$=\frac{\Lambda_1^{j-i-1}}{C_1}[f-(1-f)\phi_1^*](1+\phi_1)+\frac{\Lambda_2^{j-i-1}}{C_2}[f-(1-f)\phi_2^*](1+\phi_2). \tag{A27}$$

Finally,

$$\langle S_{34}\rangle = \sum_{\theta_i}\sum_{\theta_{i+1}}\cdots\sum_{\theta_{j-1}}\sum_{\theta_j}p_s(\theta_i)p(\theta_{i+1}|\theta_i)p(\theta_{i+2}|\theta_{i+1})\cdots p(\theta_j|\theta_{j-1})\theta_i$$

$$\times\exp(\mu\theta_{i+1})\exp(\mu\theta_{i+2})\cdots\exp(\mu\theta_{j-1})\theta_j, \tag{A28}$$

which in matrix notation is given by

$$\langle S_{34}\rangle = (1 \quad -1)\underbrace{\begin{pmatrix}p_{AA}&p_{AB}\\p_{BA}&p_{BB}\end{pmatrix}\begin{pmatrix}e^{\mu}&0\\0&e^{-\mu}\end{pmatrix}\cdots\begin{pmatrix}p_{AA}&p_{AB}\\p_{BA}&p_{BB}\end{pmatrix}\begin{pmatrix}e^{\mu}&0\\0&e^{-\mu}\end{pmatrix}}_{j-i-1\ \text{factors}}\begin{pmatrix}p_{AA}&p_{AB}\\p_{BA}&p_{BB}\end{pmatrix}\begin{pmatrix}p_A\\-p_B\end{pmatrix} \tag{A29}$$

$$=(1-1)\left[\frac{1}{C_1}\mathbf{A}_1\Lambda_1^{j-i-1}+\frac{1}{C_2}\mathbf{A}_2\Lambda_2^{j-i-1}\right]\begin{pmatrix}p_{AA}p_A-p_{AB}p_B\\p_{BA}p_A-p_{BB}p_B\end{pmatrix}$$

$$=\frac{\Lambda_1^{j-i-1}}{C_1}(1+\phi_1)[fp_{AA}-(1-f)p_{AB}-\phi_1^*(fp_{BA}-(1-f)p_{BB})]$$

$$+\frac{\Lambda_2^{j-i-1}}{C_2}(1+\phi_2)[fp_{AA}-(1-f)p_{AB}-\phi_2^*(fp_{BA}-(1-f)p_{BB})]. \tag{A30}$$

[1] V. S. Pande, A. Y. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000), and references therein.

[2] O. Kratky and G. Porod, Recl. Trav. Chim. Pays-Bas. **68**, 1106 (1949).

[3] N. Saitô, K. Takahashi, and Y. Yunoki, J. Phys. Soc. Jpn. **22**, 219 (1967).

[4] R. A. Harris and J. E. Hearst, J. Chem. Phys. **44**, 2595 (1966); M. G. Bawendi and K. F. Freed, J. Chem. Phys. **83**, 2491 (1985); J. B. Lagowski and J. Noolandi, J. Chem. Phys. **95**, 1266 (1991); L. Harnau, R. G. Winkler, and P. Reineker, J. Chem. Phys. **102**, 7750 (1995).

[5] Y. Muroga, Macromolecules **21**, 2751 (1988); Y. Muroga, H. Tagawa, Y. Hiragi, T. Ueki, M. Kataoka, Y. Izumi, and Y. Amemiya, *ibid.* **21**, 2756 (1988).

[6] G. H. Fredrickson, Macromolecules **22**, 2746 (1989).

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

[8] D. Bensimon, D. Dohmi, and M. Mezard, Europhys. Lett. **42**, 97 (1998); A. Buhot and A. Halperin, Phys. Rev. Lett. **84**, 2160 (2000); M. N.

Tamashiro and P. Pincus, Phys. Rev. E **63**, 021909 (2001).

[9] L. Gutman and A. Chakraborty, J. Chem. Phys. **101**, 10074 (1994).

[10] K. F. Freed, Adv. Chem. Phys. **22**, 1 (1972); K. F. Freed, *Renormalization Group Theory of Macromolecules* (Wiley, New York, 1987).

[11] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists* (Academic, San Diego, CA, 1995).

[12] N. T. J. Bailey, *Elements of Stochastic Processes* (Wiley, New York, 1964).