International Journal of Theoretical Physics, Vol. 26, No. 5, 1987

Statistics of the Ensemble of Primary Structures for Inhomogeneous Polymer Chains

Ariel Fernández[1]

Received January 25, 1987

We consider linear polymer chains of fixed length N made up randomly of two different kinds of monomers X and Y near the tricritical theta point. Two kinds of short-range repulsive volume-excluded interactions, X-X and Y-Y are considered. In addition, short-range attractive two-body X-Y interactions and three-body interactions are included in the theoretical treatment. We derive the distribution of the average size of the random coil over the ensemble of primary structures or X-Y sequences. The variance of thermal or configurational averages for the different sequences is calculated making use of field-theoretic methods.

1. INTRODUCTORY REMARKS AND DEFINITIONS

Linear polymer sequences of fixed length N, randomly constructed from two kinds of monomers, X and Y, have been considered in the context of generation of prebiotic information (Eigen and Schuster, 1977).

The problem of studying the distribution of thermal or statistical averages of these chains over an ensemble of randomly constructed sequences is analogous to the averaging procedure for glassy, disordered systems that arises in solid state physics (Anderson, 1983).

This analogy has been taken advantage of by mapping the set of patterns of intrachain two-body interactions for a fixed primary structure onto a spin glass, assigning a spin value to each kind of monomer; for example, $S=+1$ for X and $S=-1$ for Y.

The statistics of this quenched disordered system is relevant to the problem of finding the distribution of configuration averages in linear tRNA chains for the different *a priori* G-C sequences (Eigen and Schuster, 1977).

In order to specify the theoretical framework, we introduce the following definitions:

---

[1] Frick Laboratory, Princeton University, Princeton, New Jersey 08544.

0020-7748/87/0500-0489$05.00/0 © 1987 Plenum Publishing Corporation

Let $r(x)$ denote a chain conformation with x a contour variable indicating the position on the chain; x belongs to the interval $[0, N]$. We construct an equivalence relation as follows: $r_1 \sim r_2$ if both conformations belong to the same configuration, that is, they have the same pattern of short-range two-body interactions. Let $F = F(r)$ be a functional that respects the equivalence relation, which means

$$r_1 \sim r_2 \rightarrow F(r_1) = F(r_2) \tag{1}$$

Two kinds of averages can be defined:

(a) The thermal or statistical average for a given sequence, denoted $\bar{F}$, which is the average over all possible configurations, that is, over all r_k such that $r_i \not\sim r_j$ for $i \neq j$.

(b) The average over the ensemble of primary structures or random sequences, denoted $\langle \bar{F} \rangle$.

The two-body interactions are determined by the bare (unnormalized) coupling variable $v(x, x')$. Thus, v can take three different values corresponding to the three possible interactions between the monomer in position x and that in position x'. The interactions are X-X, Y-Y, and X-Y. The first two kinds of interactions are familiar from homogeneous polymer statistics; they represent short-range, repulsive, volume-excluded interactions (Oono and Freed, 1981).

In order to define v, we introduce a discrete stochastic process (x is here regarded as a natural number):

$$s(x) = \begin{cases} +1 & \text{if the } x\text{th monomer in the chain is X} \\ -1 & \text{otherwise} \end{cases} \tag{2}$$

The probability for $s(x)$ to take any one of the two possible values is 1/2. Each realization of $s(x)$ corresponds to a definite primary structure or X-Y sequence.

The continuous counterpart of such process is determined by the equations

$$\langle s(x) s(x') \rangle = \delta(x - x'), \qquad \langle s(x) \rangle = 0 \tag{3}$$

where the angle brackets denote an average over all possible primary structures, that is, over all sequences of N monomers. Then, $v(x, x')$ is defined as follows:

$$v(x, x') = v + A[s(x) + s(x')] + Bs(x)s(x') \tag{4}$$

where A and B are bare constants and v is the Flory bare coupling constant. Therefore

$$\langle v(x, x') \rangle = v + B\delta(x - x') \tag{5}$$

The situation of interest to us occurs in a neighbourhood of the tricritical theta point, that is, when the two-body interactions averaged over all possible realizations of $s(x)$ are almost zero, that is, $v \approx 0$.

At the theta point we can define the concept of theta-secondary structure or theta-configuration $\bar{r}_0$ containing the conformation r_0, as the set of conformations r such that $r \sim r_0$, that is,

$$\int_0^N \int_0^N dx\, dx'\, v(x, x')\, \delta(r(x) - r(x'))$$

$$= I = \int_0^N \int_0^N dx\, dx'\, v(x, x')\, \delta(r(x) - r(x'))\, \delta(r_0(x) - r_0(x')) \qquad (6)$$

where I is the two-body contribution to the potential energy and is given by

$$I = \int_0^N \int_0^N dx\, dx'\, v(x, x')\, \delta(r_0(x) - r_0(x')) \qquad (7)$$

We shall calculate dispersions over the quenched randomness making use of the field theory formalism. This requires that we introduce the contribution of a virtual field $f(r)$ in the Hamiltonian and consider the limit $f \to 0$ in the Frechet derivatives of the partition function with respect to f. The partition function is a functional that depends on the specific realization of $s(x)$ and on the field f.

2. FIELD-THEORETIC CALCULATION OF THE VARIANCE FOR CONFIGURATIONAL AVERAGES

Formally, the partition function can be expressed by means of the following integral:

$$Q = \int_{r(0)=0,\, r(N)=R} dm(r(x)) \exp[-H(r(x))] \qquad (8)$$

where $m(r(x))$ is a measure in the conformation space and the Hamiltonian H is a functional of the path $r(x)$. This Hamiltonian contains the short-range, volume-excluded interactions (Kholodenko and Freed, 1983), the remaining two-body interactions, the three-body interactions and the contribution from the virtual field.

Instead of considering the integration over all possible paths, we discretize the range of the contour variable. This is done by introducing an arbitrary partition P_M given by

$$P_M = \{r_0, r_1, r_2, \ldots, r_M\}, \qquad \text{where} \quad r_j = r(jN/M) \qquad (9)$$

Thus, with respect to P_M, the expression (8) can be replaced by a multiple integration over $M+1$ variables so that

$$Q = \lim_{M\to\infty} \int \prod_{j=0}^{M} dr_j\, \delta(r_0)\, \delta(r_M - R) \exp(-H_M) = \lim_{M\to\infty} Q_M \tag{10}$$

where H_M is a function of $M+1$ variables instead of being a functional; this function is defined as follows:

$$H_M = \frac{1}{2} \sum_{i=0}^{M} \left(\frac{r_{i+1} - r_i}{N/M}\right)^2 + H'_M = H^e_M + H'_M \tag{11}$$

The term H'_M is functionally dependent on $f(R)$ and contains the two-body interaction term, parametrically dependent on v, A, and B, and the three-body interaction term, parametrically dependent on the Flory bare three-body interaction constant C:

$$H'_M = \frac{1}{2}\left[\sum_{i,j=0}^{M} v(iN/M, jN/M)\, \delta_{r_i r_j} + \frac{1}{3} C \sum_{k=0}^{M} \delta_{r_i r_j}\, \delta_{r_i r_k}\right] - f(r_M) \tag{12}$$

The Kronecker symbol $\delta_{r_i r_j}$ is defined as follows:

$$\delta_{r_i r_j} = \begin{cases} +1 & \textit{if } r_i = r_j \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

We now consider the size R of the random coil, $R(r(x)) = r_M$. Throughout what follows, a P_M-discretization of the problem with $M \gg N$ will be adopted.

The Green's function G for a disordered linear polymer is given by

$$G(R) = \left.\frac{\partial \ln Q}{\partial f(R)}\right|_{f=0} = \lim_{M\to\infty} \left.\frac{\partial \ln Q_M}{\partial f(R)}\right|_{f=0} \tag{14}$$

We now concentrate on the calculation of the variance of the thermal average $\overline{R^2}$. This quantity is given by

$$\overline{R^2} = -\left.\frac{\partial^2 g(p)}{\partial p^2}\right|_{p=0} \tag{15}$$

where $g(p)$ is the Fourier transform of the Green's function in momentum space.

The central problem is to evaluate the variance

$$V(\overline{R^2}) = \langle(\overline{R^2} - \langle[\overline{R^2}]\rangle)^2\rangle = V \tag{16}$$

$$V = \frac{1}{L} \int\!\!\int R_1^2 R_2^2 \left.\frac{\partial^2 \ln\langle Q(L)\rangle}{\partial f(R_1)\, \partial f(R_2)}\right|_{f=0} dR_1\, dR_2 \tag{17}$$

(L = large integer), where $Q(L)$ is a partition function for L identical copies of a single polymer chain corresponding to a single primary structure.

Therefore the problem is reduced to the computation of the average of $Q(L)$ over the ensemble of all possible realizations $s(x)$.

We define the effective partition functions $Q_{\text{eff}}(\alpha)$ corresponding to homogeneous polymer chains of length N in such a way so that the following relation holds:

$$\langle Q(L)\rangle = \prod_{\alpha=1}^{L} Q_{\text{eff}}^{(\alpha)} \tag{18}$$

The effect of the existence of a quenched randomness determining distinctive sequences is to enhance the repulsiveness of volume-excluded interactions and to diminish the effect of three-body interactions. In order to prove this, we consider the average

$$\langle Q(L)\rangle = \lim_{M\to\infty} \int \prod_{\alpha=1}^{L} \prod_{i=0}^{M} dr_{\alpha i}\, \delta(r_{\alpha 0})\, \delta(r_{\alpha M} - R_\alpha) \exp[-H_M(L)] \tag{19}$$

where

$$\begin{aligned} H_M(L) = \sum_{\alpha=1}^{L} H_M^e(\alpha) + \sum_{\alpha=1}^{L} \sum_{i,j=0}^{M} \Bigg[\frac{v}{2}\, \delta_{r_{\alpha i} r_{\alpha j}} & \\ -\frac{1}{2} A^2 \sum_{\beta\neq\alpha} \sum_{k=0}^{M} \Bigg(\delta_{r_{\alpha i} r_{\alpha j}} \delta_{r_{\beta i} r_{\beta k}} & \\ -\frac{C-3A^2}{3A^2}\, \delta_{r_{\alpha i} r_{\alpha j}} \delta_{r_{\alpha i} r_{\alpha k}} \Bigg) - f(r_{\alpha M}) \Bigg] & \end{aligned} \tag{20}$$

Therefore, the effective Hamiltonian is given by

$$H_{\text{eff}\, M} = H_M^e + \sum_{i,j=0}^{M} \left(\frac{v_{\text{eff}}}{2} \delta_{r_i r_j} + \frac{C_{\text{eff}}}{6} \sum_{k=0}^{M} \delta_{r_i r_j} \delta_{r_j r_k} \right) - f(r_M) \tag{21}$$

The two-body effective interactions are of the short-range, volume-excluded type with a bare coupling constant

$$v_{\text{eff}} = v + 2A^2 \lim_{M\to\infty} \sum_{j=0}^{M} \frac{\partial \ln[\prod_{\alpha=1}^{L} Q_j(\alpha)]}{\partial f(R_\alpha = R = 0)} \tag{22}$$

where $Q_j(\alpha)$ is the effective partition function for a homogeneous linear chain of length jN/M such that the two endpoints are coincident.

The effective three-body interaction constant is

$$C_{\text{eff}} = C - 3A^2 \tag{23}$$

We evaluate the variance to first order making use of equations (17)–(23) [for a review of the perturbation methods implemented in homogeneous polymer statistics, see Kholodenko and Freed (1983)]; such a calculation gives

$$V=(1/128)N^2A^2(\pi^4+4\pi^3) \tag{24}$$

The configurational averages are more spread as the amount of stored information given by the length of the chains increases (Eigen and Schuster, 1977; Anderson, 1983). The first-order perturbative calculation reveals that the limit of vanishing three-body interactions is a well-defined limit in a neighborhood of the theta point, since the variance is independent of the bare Flory constant C.

REFERENCES

Anderson, P. W. (1983). *Proceedings of the National Academy of Sciences of the USA*, **80**, 3380.

Eigen, M., and Schuster, P. (1977). *Naturwissenschaften*, **64**, 541.

Kholodenko, A. L., and Freed, K. F. (1983). *Journal of Chemical Physics*, **78**, 7390.

Oono, Y., and Freed, K. F. (1981). *Journal of Chemical Physics*, **75**, 993.