Biophysical Chemistry, 32 (1988) 167–171
Elsevier

BPC 01321

Dispersion of tertiary structures for an ensemble of primary sequences at an externally induced transition of correlation regimes

Ariel Fernández

Max-Planck-Institut für Biophysikalische Chemie, Am Faßberg, 3400 Göttingen, F.R.G.

Received 4 July 1988
Accepted 6 September 1988

Statistical mechanics; Random inhomogeneous polymer; Quenched disorder; Replica trick;
Polymer phase transition; Coil adsorption

We consider the denaturation of globular regimes for an ensemble of random polymer sequences with long-range interactions, constituting a quenched disorder. The sensitivity of thermally averaged configurational correlations to primary sequences is shown to reach a maximum at the coil-globule transition.

1. Introduction

Major effort has been devoted to the inference of accessible tertiary structures for biopolymers with a given primary sequence subject to environmental constraints [1–4].

We shall consider a weaker problem which is susceptible to a statistical-mechanical treatment, namely, what are the configurational statistics of an ensemble of randomly constructed polymer sequences when the folding is externally induced by an attractive field? The external field which induces a phase transition requires further specification [5,6]. We shall assume that it is localized and determined by the adsorption of the coil by particles (vesicles) or interphases whose dimensions do not interfere with the measurement of the globular size by light scattering.

Since the domain of interest to us is the good solvent region we need to include volume-excluded long-range interactions for a long flexible chain [7]. This complicates the situation in a non-trivial way, since the Markovian beads-on-the-string model with quenched disorder does not hold [4]. The bending is non-Markovian.

We shall resort to the standard conformational space-functional integral formalism proposed by Edwards [8]. The Hamiltonian for the disordered chain is made up of five contributions:

$$H(x(t)) = H^{e} + H^{(2)} + H^{(3)} + H^{F} + H^{f} \tag{1}$$

$$H^{e} = \frac{1}{2}\int_0^N \mathrm{d}t \left(\left\| \frac{\partial x(t)}{\partial t} \right\| \right)^2 \tag{2}$$

$$H^{(2)} = \frac{1}{2}\int_0^N \int_0^N \mathrm{d}t\, \mathrm{d}t' v(t,t') \delta(\boldsymbol{x}(t) - \boldsymbol{x}(t'))$$

= two-body volume-excluded interactions (3)

$$H^{(3)} = \frac{C}{3!}\int_0^N \int_0^N \int_0^N \delta(\boldsymbol{x}(t) - \boldsymbol{x}(t')) \times \delta(\boldsymbol{x}(t) - \boldsymbol{x}(t''))\, \mathrm{d}t\, \mathrm{d}t'\, \mathrm{d}t''$$

= three-body interactions (4)

$$H^{F} = \int_0^N (F_t(\boldsymbol{x}(t))/T)\, \mathrm{d}t$$

= attractive external field term (5)

$$H^{f} = \int \mathrm{d}\boldsymbol{R} f(\boldsymbol{R}) \delta(\boldsymbol{x}(N) - \boldsymbol{R})$$

= virtual field contribution (6)

Correspondence address: A. Fernández, Max-Planck-Institut für Biophysikalische Chemie, Am Faßberg, 3400 Göttingen, F.R.G.

Here, the 't' is the contour variable indicating the position on the chain, $x(t)$ denotes the position vector of the chain segment located between t and $t+\delta t$, $v(t,t')$ is the bare two-body volume-excluded coupling constant, C the bare Flory interaction constant for the three-body contribution, $F_t(x(t))$ the external field exerted at position t on the chain and $f(\boldsymbol{R})$ the virtual field controlling the endpoint of the chain (the only relevant case consists of taking the limit $f=0$).

Obviously the bare Hamiltonian depends on the primary sequence. For simplicity, we shall consider the case of two different kinds of monomers only. The strands are assumed to be randomly constructed. Thus, a specific primary sequence is determined unambiguously by a specific realization of the random variable $s(t)$, defined by:

$$\langle s(t)\rangle = 0 \tag{7}$$

and

$$\langle s(t)s(t')\rangle = \delta(t-t') \tag{8}$$

Where $\langle \quad \rangle$ denotes average over the ensemble of primary structures and we have assumed that $s(t)$ can only take the values ± 1 depending on the kind of monomer at position t.

The Hamiltonian depends on the specific realization of $s(t)$. Consequently, we define

$$v(t,t') = v + A(s(t)+s(t')) + Bs(t)s(t') \tag{9}$$

where A and B are unrenormalized constants. A localized field representing the trapping of the coil by the adsorption well is given by:

$$\frac{F_t(x(t))}{T} = -\ln\{1 + T^{-1}(F_{\mathrm{eff}} + \Lambda s(t))\delta(x(t))\} \tag{10}$$

That is, the field is localized at $x=0$. The constants F_{eff} and Λ have the dimensions of temperature. The external field must obviously be attractive in order to induce a phase transition. As the temperature is raised beyond a certain critical value T_c, the well F_t/T no longer holds the chain and a different correlation range regime (CRR), the coil, emerges.

First-order corrections to the bare coupling in the good solvent region, near the theta point ($|v| < N^{-1/2}$), are given by:

$$\delta v(t,t') = v(t,t')\int_0^N\int_0^N \mathrm{d}t_1\,\mathrm{d}t_2 v(t_1,t_2) \times \int G_{tt_2}(\boldsymbol{p})G_{t't_1}(\boldsymbol{p})\frac{d^3\boldsymbol{p}}{8\pi^3} \tag{11}$$

where $G_{t_at_b}$ is the Gaussian correlator given in momentum representation by

$$G_{t_at_b} = \exp(-\tfrac{1}{2}p^2|t_a-t_b|) \tag{12}$$

The fixed-end partition function is

$$Z(N) = \int_{x(0)=0} \mathrm{d}m(x(t)) \exp(-H(x(t)) \tag{13}$$

where $m(x(t))$ is a measure in the space of paths. Thus, the Green function $G(N,\boldsymbol{R})$ is given by:

$$\left.\frac{\partial Z(N)}{\partial f(\boldsymbol{R})}\right|_{f=0} = G(N,\boldsymbol{R}) = \overline{\delta(x(N)-\boldsymbol{R})} \tag{14}$$

where the bar above a quantity indicates average over all conformations for a given configuration (thermal average).

The next stage is to introduce an effective homopolymer so that the coil-globule transition belongs to the same universality class as that corresponding to the quenched disorder. This is carried out making use of the replica trick for averaging over the ensemble of primary structures [9,10].

2. Quenched disorder and the replica trick

We shall denote by $Z(N)^M$, where M is a positive integer, the partition function for M identical replicas of an arbitrarily chosen strand. This function can be extended analytically for M real [9]. The following relation will prove very useful:

$$\langle \ln Z(N)\rangle = \lim_{M\to 0}\frac{1}{M}(\langle Z(N)^M\rangle - 1) \tag{15}$$

The replica symmetry breaking thus appears in the computation of $\langle Z(N)^M\rangle$. Let α and β be dummy indices labelling replicas, then we have:

$$\langle Z(N)^M\rangle = \prod_{\beta=1}^{M} Z_{\mathrm{eff},\beta}(N) \tag{16}$$

where $Z_{\mathrm{eff},\beta}$ is the partition function for an effective Hamiltonian corresponding to a homopolymer, i.e., it contains no disorder. This Hamiltonian, however, couples the replicas labelled by the M values of the dummy index. The effective partition function can be evaluated by means of a perturbative expansion as follows:

$$Z_{\mathrm{eff}} = Z^{(0)} + Z_{\mathrm{eff}}^{(2)} + Z_{\mathrm{eff}}^{(3)} + \dots \tag{17}$$

where $Z^{(0)}$ is the partition function for a homogeneous polymer strand with Hamiltonian: $H^{(0)} = H^{\mathrm{e}} + H_{\mathrm{eff}}^{\mathrm{F}} + H^{\mathrm{f}}$. The term $H_{\mathrm{eff}}^{\mathrm{F}}$ represents the coupling to a field which contains no disorder, i.e., with $\Lambda = 0$. The perturbation terms are given by:

$$\begin{aligned} Z_{\mathrm{eff}}^{(2)} = & -\int_0^{\boldsymbol{R}} \mathrm{d}\boldsymbol{x} \int_0^{N} \mathrm{d}t'' \int_0^{t''} \mathrm{d}t' \int_0^{t'} \mathrm{d}t \Big\{ v - A^2 \sum_{\beta\neq\alpha} G_{\beta 0} \\ & \times (t',\boldsymbol{x}) G_{\beta 0}(t''-t',\boldsymbol{0}) G_{\beta 0} \\ & \times (N-t'',\boldsymbol{R}-\boldsymbol{x}) \Big\} \\ & \times G_{\alpha 0}(t,\boldsymbol{x}) G_{\alpha 0}(t'-t,\boldsymbol{0}) G_{\alpha 0} \\ & \times (N-t',\boldsymbol{R}-\boldsymbol{x}) \end{aligned} \tag{18}$$

Where G_0 is the propagator with elastic Hamiltonian.

In momentum representation:

$$G_0(s,\boldsymbol{R}) = (2\pi)^{-3} \int d^3p \frac{\exp(i\boldsymbol{p}\boldsymbol{R})}{\boldsymbol{p}^2/2 + s} \tag{19}$$

we also have:

$$\begin{aligned} Z_{\mathrm{eff}}^{(3)} = & -(C - 3A^2) \int_0^{\boldsymbol{R}} \mathrm{d}\boldsymbol{x} \int_0^{N} \mathrm{d}t'' \int_0^{t''} \mathrm{d}t' \int_0^{t'} \mathrm{d}t G_0 \\ & \times (t,\boldsymbol{x}) G_0(t'-t,\boldsymbol{0}) G_0(t''-t',\boldsymbol{0}) G_0 \\ & \times (N-t'',\boldsymbol{R}-\boldsymbol{x}) \end{aligned} \tag{20}$$

Thus, the effective Hamiltonian reads:

$$\begin{aligned} H_{\mathrm{eff}} = & H^{\mathrm{e}} + H_{\mathrm{eff}}^{\mathrm{F}} + H^{\mathrm{f}} \\ & + \frac{v_{\mathrm{eff}}}{2} \int_0^N \int_0^N \mathrm{d}t\, \mathrm{d}t' \delta(\boldsymbol{x}(t) - \boldsymbol{x}(t')) \\ & + \frac{C_{\mathrm{eff}}}{3!} \int_0^N \int_0^N \int_0^N \mathrm{d}t\, \mathrm{d}t'\, \mathrm{d}t'' \delta(\boldsymbol{x}(t) \\ & - \boldsymbol{x}(t')) \delta(\boldsymbol{x}(t) - \boldsymbol{x}(t'')) \end{aligned} \tag{21}$$

where:

$$C_{\mathrm{eff}} = C - 3A^2 \tag{22}$$

and

$$v_{\mathrm{eff}} = v + 2A^2 \lim_{M\to 0} \int_0^N \frac{\partial \ln \tilde{Z}_t^M}{\partial f(\boldsymbol{R}_\alpha)} \Bigg|_{\substack{f=0 \\ \boldsymbol{R}_\alpha = 0}} \mathrm{d}t \tag{22'}$$

The function $\tilde{Z}_t^M$ for integer M is the partition function averaged over the ensemble of primary structures for M replicas of a strand of length t whose endpoints coincide.

In order to determine the CRR for the effective Hamiltonian, we need to introduce a different representation. We start by discretizing the domain of the contour variable introducing a partition:

$$(j = 1, 2, \dots, N);\ \boldsymbol{x}(j) = \boldsymbol{x}_j \tag{23}$$

We also introduce transition operators $\hat{W}_j$ defined by:

$$\hat{W}_j G(j-1,\boldsymbol{x}_{j-1}) = G(j,\boldsymbol{x}_j) \tag{24}$$

Thus, for $j = 1, 2, 3, 4$, we have:

$$\begin{aligned} & \hat{W}_j G(j-1,\boldsymbol{x}_{j-1}) \\ & \quad = \exp\big(-\bar{F}(\boldsymbol{x}_j)/T\big) \\ & \qquad \times \int \mathrm{d}\boldsymbol{x}_{j-1} g(\boldsymbol{x}_j - \boldsymbol{x}_{j-1}) G(j-1,\boldsymbol{x}_{j-1}) \end{aligned} \tag{25}$$

where $\bar{F}$ is a field with $\Lambda = 0$ and the coupling between links is assumed to be identical:

$$g(\boldsymbol{x}_j - \boldsymbol{x}_{j-1}) = (4a^2)^{3/2} \exp\left(-\frac{1}{4a^2} (\boldsymbol{x}_j - \boldsymbol{x}_{j-1})^2 \right). \tag{26}$$

Here a is the average distance between adjacent links. For $j \geq 4$, the transition operators are different, since we must consider long-range interactions of the repulsive type which we denote $\hat{\hat{W}}_j$):

$$\hat{\hat{W}}_j G(j-1,\boldsymbol{x}_{j-1})$$

$$= \exp\left(-\bar{F}(\boldsymbol{x}_j)/T\right) \int \mathrm{d}\boldsymbol{x}_{j-1} g(\boldsymbol{x}_j - \boldsymbol{x}_{j-1})$$

$$\times \int \prod_{\tau=4}^{j-1} \left\{ \Omega^{(2)}(\boldsymbol{x}_j, \boldsymbol{x}_{j-\tau}) \right.$$

$$\left. \times G(j-\tau, \boldsymbol{x}_{j-\tau})\, \mathrm{d}x_{j-\tau} \right\}$$

$$\times \prod_{\tau,\tau'=4}^{j-1} \left\{ \Omega^{(3)}(\boldsymbol{x}_j, \boldsymbol{x}_{j-\tau}, \boldsymbol{x}_{j-\tau'}) G(j-\tau, \boldsymbol{x}_{j-\tau}) \right.$$

$$\times G(j-\tau', \boldsymbol{x}_{j-\tau'})$$

$$\left. \times \mathrm{d}\boldsymbol{x}_{j-\tau}\, \mathrm{d}\boldsymbol{x}_{j-\tau'} \right\} G(j-1, \boldsymbol{x}_{j-1}) \tag{27}$$

For clarification, eq. 27 can be rewritten as:

$$\hat{\hat{W}}_j G(j-1,\boldsymbol{x}_{j-1}) = \hat{W}_j \hat{\Omega}^{(2)} \hat{\Omega}^{(3)} G(j-1, \boldsymbol{x}_{j-1}) \tag{28}$$

where $\hat{\Omega}^{(2)}$ and $\hat{\Omega}^{(3)}$ are the two- and three-body interaction transition operators, respectively. Their kernels are given by:

$$\Omega^{(2)}(\boldsymbol{x}_j, \boldsymbol{x}_{j-\tau}) = \exp\left\{ -\frac{v_{\mathrm{eff}}}{2} \delta_{x_j x_{j-\tau}} \right\} \tag{29}$$

and

$$\Omega^{(3)}(\boldsymbol{x}_j, \boldsymbol{x}_{j-\tau}, \boldsymbol{x}_{j-\tau'}) = \exp\left\{ -\frac{C_{\mathrm{eff}}}{3!} \delta_{x_j x_{j-\tau}} \delta_{x_j x_{j-\tau'}} \right\} \tag{30}$$

We are left with the diffusion equations, eqs. 25 and 27. However, the functions $G(j,\boldsymbol{x}_j)$ do not form a Markov chain, since the operator $\hat{\hat{W}}_j$ ($j > 4$) and the 'vector' $G(j-1,\boldsymbol{x}_{j-1})$ are not statistically independent.

The globular state is characterized by the spectrum of $\hat{\hat{W}}_N$ for N large. The signature for that specific CRR is that the largest real eigenvalue, λ, is not an accumulation point in the spectrum but represents a discrete level. The temperature T_c at which the phase transition between the two CRRs occurs is the temperature at which the level λ splits off from the continuum part of the spectrum. The situation of discrete largest eigenvalue is preserved for $T < T_c$. Since the term corresponding to the largest eigenvalue dominates the Green function, the free energy $\Delta F(N,R)$ for the effective homopolymer in the globular CRR is given by:

$$\Delta F(N,R) = -T \ln \lambda^N = -TN \ln \lambda \tag{31}$$

Thus, the square end-to-end distance averaged over all conformations (thermal average) for the globular CRR is:

$$\overline{R^2} = \frac{\int R^2 \exp(-\Delta F(N,R)/T)\, \mathrm{d}R}{Z(N)} \tag{32}$$

whereas for the random coil CRR it is given by

$$\overline{R^2} = -\left. \frac{\partial^2 G(N,p)}{\partial p^2} \right|_{p=0} . \tag{33}$$

where $g(N,p)$ is the Fourier-transform of $G(N,R)$. The dispersion relation (variance) for $\overline{R^2}$ over the ensemble of primary structures, $V(\overline{R^2})$, is a quantity of the utmost importance in the design of light-scattering experiments in order to establish the sensitivity of the configuration to the primary

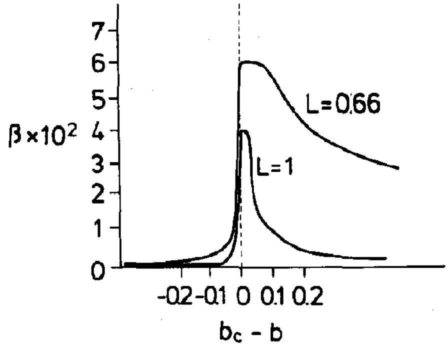

Fig. 1. Calculation of the variance for $\overline{R^2}$ over the ensemble of primary sequences. Choice of parameters: $a^2 = 100$ Å^2 = mean square length of a link; $N = L \times 10^3$; $u = \frac{1}{2}N^{-1/2}$; $A = \frac{1}{4}N^{-1/2}$; $B = N^{-1}$. External field: $F_{\mathrm{eff}} = 270$ K, $\Lambda = 10$ K, $T_c = 340$ K. $b_c - b = (T_c - T)/T$. $V(\overline{R^2})^{1/2}/\langle\overline{R^2}\rangle = \beta$.

sequence. In a neighborhood of the phase transition we have:

$$V\left(\overline{R^2}\right) = \left\langle \left(\overline{R^2} - \langle \overline{R^2} \rangle\right)^2 \right\rangle$$
$$= \lim_{M \to 0} \iint \mathrm{d}R_\alpha \, \mathrm{d}R_\beta R_\alpha^2 R_\beta^2 \times \left. \frac{\partial^2 \ln(\langle Z^M \rangle - 1)}{\partial f(R_\alpha) \partial f(R_\beta)} \right|_{f=0} \tag{34}$$

A perturbative analysis gives for the random coil regime:

$$V\left(\overline{R^2}\right) = \frac{N^2 A^2 (4+\pi)}{128\pi^3} \tag{35}$$

A numerical computation for the dispersion is presented in fig. 1. It holds for an entire neighborhood of the phase transition. Upon direct inspection, we can see that maximum sensitivity is achieved at the phase transition point and that the globular CRR presents a much higher degree of dependence on the primary structure. This confirms the relevance of the concept of the globular state as introduced by Lifshitz [1,4] for the problem of biopolymer folding. A simple computation for $L = 1$ gives, at $b_c - b = 0.2$, a net stability for the effective globule of about 15 kcal/mol with respect to the coil at the same reduced temperature.

References

1 I.M. Lifshitz, Zh. Eksp. Teor. Fiz. 65 (1973) 1100 (Sov. Phys. JETP 38 (1974) 545).
2 T.E. Creighton, Proteins, structure and molecular properties (Freeman, San Francisco, 1983).
3 M. Eigen and P. Schuster, The hypercycle (Springer, Berlin, 1979) ch. 1.
4 A.Yu. Grosberg and E.I. Shakhnovich, Zh. Eksp. Teor. Fiz. 91 (1986) 2159 (Sov. Phys. JETP 64 (1986) 1284).
5 P.G. de Gennes, J. Phys. (Paris) 37 (1976) 1445.
6 M. Lax, Chem. Phys. 10 (1975) 285.
7 A.L. Kholodenko and K.F. Freed, J. Chem. Phys. 78 (1983) 7390.
8 S.F. Edwards, Proc. Roy. Soc. Lond. 85 (1965) 613.
9 A. Fernández and H. Rabitz, Biophys. Chem. 28 (1987) 89.
10 A. Fernández, Z. Phys. Chem. (Leipzig) (1988) in the press; A. Fernández, Ber. Bunsenges. Phys. Chem. 91 (1987) 753.