Theory of phase transitions of the coil-globule type in a heteropolymer chain with disordered sequence of links

A. Yu. Grosberg and E. I. Shakhnovich

Institute of Chemical Physics, USSR Academy of Sciences
(Submitted 8 May 1986)
Zh. Eksp. Teor. Fiz. **91**, 2159–2170 (December 1986)

A coil-globule phase transition (collapse) in a heteropolymer chain with random sequence of links of various types under the influence of external fields (which are different for the different types of link) is considered. It is shown that the system is renormalizable. A renormalization transformation is constructed, which unifies neighboring links into blocks and leads to a chain with a new linear structure. The equation for renormalization of the probability distributions of the effective external fields acting on block links is obtained in explicit form and solved. In most cases of physical interest, introduction of disorder in a sequence of links leaves the globule-coil phase transition in the same universality class to which it belongs for the homopolymer. It is shown that the swelling of the globule near the transition point is characterized by a single scale that is independent of the sequence of the links. Increase of the distance from the point of transition into the globular region enhances the dependence of the spatial structure and of the free energy of the globule on the sequence of links. The mean values and variances of the free energy and of the dimension of the random chain in the phase-transition region are calculated.

1. INTRODUCTION

This study was prompted by the demands of protein physics and continues a trend initiated by I. M. Lifshitz's classical paper.[1] In the physical context, one of the important unsolved problems in this field is the need to go beyond the scope of the model used for a homogeneous homopolymer in the theory of coil-globule phase transitions,[2] and the need to study the picture of collapse and structure of condensed phases for macromolecules of heteropolymers with inhomogeneous disordered link sequence (primary structure). The importance of this problem for biopolymers is obvious, and its formulation was discussed already in Ref. 1.

Our present purpose is to analyze the role of an inhomogeneous sequence, frozen in the course of chain formation, of links in the picture of the simplest transition of the coil-globule type. To be specific, we consider the collapse of an ideal (immaterial) heteropolymer under the influence of an external field. This model system is of independent physical interest in connection with polymer adsorption (see the bibliography in Ref. 4) and the helix-coil transition in DNA.[5] Our results are of primary interest, however, in connection with the question of conformational entropy of a heteropolymer in a spatially inhomogeneous state, inasmuch as in the framework of I. M. Lifshitz's approach[1] this question is central in the development of a consistent theory of heteropolymer globules with volume interaction (see Ref. 6 for details; we call attention also to Obukhov's alternate approach[7]).

2. PRINCIPAL DEFINITIONS AND TERMINOLOGY

The partition function of an immaterial chain with fixed ends, having the meaning of a Green's function, can be written in the form

$$G_t(\mathbf{x}) = \left[\prod_{\tau=1}^{t-1} \hat{Q}_\tau \right] \delta(\mathbf{x}), \tag{1}$$

where the transition operators $\hat{Q}_\tau$ are given by the expression[1-3]

$$\hat{Q}_\tau = \exp(-\varphi_\tau(\mathbf{x})/T)\hat{g}, \qquad \hat{g}\psi = \int g(|\mathbf{x}-\mathbf{x}'|)\psi(\mathbf{x}')d^d x'.$$

Here $\varphi_\tau(\mathbf{x})$ is the potential of the external field acting on link number τ; these potentials are different for different τ, since we are dealing here with a heteropolymer. In the transition-temperature vicinity of interest to us the correlation radius is large, i.e., the typical conformation of the globule includes long free loops and only a relatively small core. Accordingly, all the external fields can be regarded as localized at the point

$$\exp(-\varphi_\tau(\mathbf{x})/T) = 1 + \beta_\tau \delta(\mathbf{x}). \tag{2}$$

We assume for simplicity the couplings between the links $g(\mathbf{x} - \mathbf{x}')$ to be identical and independent of the species of the links. In the simplest example the couplings g are Gaussian, and in the general case we have for the kernel g in the k-representation

$$g_k = 1 - (ka)^2/2d, \qquad ka \ll 1, \tag{3}$$

where a is the length of the link.

Knowing the Green's function, we easily calculate the total partition function Z_t and the distribution function in the spatial points $\mathbf{x}$ of the end of the chain $\psi_t(\mathbf{x})$:

$$Z_t = \int G_t(\mathbf{x})d^d x, \qquad \psi_t(\mathbf{x}) = G_t(\mathbf{x})/Z_t.$$

We shall find useful also the following obvious recurrence relations:

$$G_{t+1}(\mathbf{x}) = \hat{Q}_t G_t, \tag{4}$$

$$\psi_{t+1}(\mathbf{x}) = \hat{Q}_t \psi_t \Big/ \int_\Delta \hat{Q}_t \psi_t \, d^d x, \tag{5}$$

$$Z_{t+1}/Z_t = \int \hat{Q}_t \psi_t \, d^d x. \tag{6}$$

For the free energy we have hence

$$F_N = -T \ln Z_N = -T \ln \prod_t Z_t/Z_{t-1} = -T \sum_t \ln \int \hat{Q}_t \psi_t \, d^d x. \tag{7}$$

We emphasize that the disordered primary structure of the chain, i.e., within the scope of our model, a sequence of operators $\hat{Q}_t$ or simply of numbers β_t, is assumed to be fixed (frozen). If this structure is known, we can calculate from (4)–(7) all the necessary quantities step by step. But if we are considering a random primary structure, then the quantities $G_t(\mathbf{x})$, Z_t, $\psi(\mathbf{x})$, F_N, etc., turn out to be random. It is shown in Ref. 8 that in this case it is most natural to describe the difference between the coil and globule states in terms of probability distributions of the quantities $\psi_t(\mathbf{x})$, i.e., $\mathscr{P}_t\{\psi(\mathbf{x})\}$. It is precisely in the globular regime that the distribution $\mathscr{P}_t\{\psi(\mathbf{x})\}$ approaches exponentially, with increase chain length t, a limit that is independent of t:

$$\mathscr{P}_t\{\psi(\mathbf{x})\} \underset{t \to \infty}{\longrightarrow} \mathscr{P}\{\psi(\mathbf{x})\} \quad \text{(for a globule)} \tag{8}$$

in analogy with the fact that in a homopolymer the quantity $\psi_t(\mathbf{x})$ itself becomes stationary with increase of t; for a coil, on the contrary, $\mathscr{P}_t(\psi)$ does not tend at all to a stationary value.

The property (8) is well known in the theory of one-dimensional disordered systems, where relations of type (4)–(6) are called Dyson–Schmidt equations. Usually $G_t(\mathbf{x})$ is regarded as a vector in Hilbert space, and $\psi_t(\mathbf{x})$ defines its "direction" or "phase." Expression (8) is referred to as the stabilization of the distribution of the phase variable. This stabilization always takes place in models of the Ising type, while the coil regime and the coil-globule transition have no analogs in these models.

The most meaningful interpretation of (8) can be given by likening t formally to time and regarding (4) as a dynamic equation of motion. In this analogy, the globule corresponds to a stochastic regime. Indeed, the various operators $\hat{Q}_t$ do not commute with one another, therefore each succeeding application of the operator in (4) leads to "rotation" of the vector $G_t(\mathbf{x})$ through a finite angle, meaning thus instability of the "motion." It is important that the instability is a property of a dynamic equation that includes noncommuting operators. The primary structure plays therefore the role not of external noise, but of a sequence of signs in the expressions for the initial conditions.[9] The foregoing pertains precisely to the globular regime, because the result of the action of the commutator of different operators $\hat{Q}_t$ on the "coil" functions ψ_t, inasmuch as they are "smeared out" over a large volume outside the localization region of the fields, turns out to be negligibly small, so that the coil corresponds to stable dynamic behavior. The coil-goluble transition is therefore none other than a transition from a dynamic to a stochastic regime in Eq. (4). Accordingly, the limiting distribution is an exact analog of a microcanonical distribution in statistical mechanics.

The Green's function (1) for $\mathbf{x} = 0$ can, with allowance for (2), represented by the series

$$G_{t'}(\mathbf{x} = 0) = \sum \underbrace{\overset{t_1 \qquad t_2 \qquad t_3}{\rule{3cm}{0.4pt}}}_{} \cdots ,$$

where each wavy line corresponds to a factor β_{t_i}, a solid segment represents the "free-field propagator" $(t - t')^{-d/2}$ and the summation is over all the t_i. The crux of the problem is the power-law form of the propagator. It shows, in particular, that allowance for the excluded volume and the topology of the loops in the fringe of the globule, if it is permissible to neglect the influence of these factors on the core of the globule, reduces simply to replacing d by a somewhat more effective quantity. We shall therefore assume d itself to be an arbitrary, not necessarily integer, quantity, subject to the only condition $d > 1$.

3. RENORMALIZATION TRANSFORMATION FOR HETEROPOLYMER

If all the external fields are pointlike (2), the basic equation (4) can be rewritten in the form

$$G_\varkappa^{(t+1)} = (1 - \varkappa^2) G_\varkappa^{(t)} + (\beta_t/a^d) \int_0^1 h_d(\varkappa) G_\varkappa^{(t)} \, d\varkappa. \tag{9}$$

To recast (4) in this form we must take its Fourier transform and use the following notation in the resultant $\mathbf{k}$-representation

$$\varkappa^2 = 1 - g_\mathbf{k}(0 \leqslant \varkappa \leqslant 1),$$
$$h_d(\varkappa) \, d\varkappa = (2\pi)^{-d} g_\mathbf{k}(2\pi^{d/2}/\Gamma(d/2)) a^d |\mathbf{k}|^{d-1} d|\mathbf{k}|.$$

According to (3) we have in the region of large distances or small $\mathbf{k}$

$$\varkappa \simeq ka(2d)^{-1/2},$$
$$h_d(\varkappa) \simeq h_d^* \varkappa^{d-1}, \quad h_d^* = (d/2\pi)^{d/2}/2\Gamma(d/2) \quad \text{for } ka \ll 1. \tag{10}$$

Equation (9) was investigated by us[10] for the special case of a homopolymer, when β_t does not depend on t, and exhibited a "critical slowing down." This property is particularly pronounced just for a heteropolymer. Indeed, the behavior of the end of the chain near the field-action point O depends mainly on the force of interaction between the point O and the last or the last few links. At the same time, the behavior of the end of the chain far from the point O is equally dependent on many links, since the end of the chain at a large distance constitutes only a multilink "tail." It can thus be stated that if $\varkappa$ is close to unity the quantity $G_\varkappa^{(t)}$ "follows up" all the high-frequency fluctuations of the random process β_t, whereas with decrease of $\varkappa$ the "temporal" behavior

of $G_\kappa^{(t)}$ is smoothed out and $G_\kappa^{(t)}$ turns out to depend on ever slower harmonics of the process β_t.

It follows from the foregoing that the heteropolymer equation (9) can be analyzed by the dynamic renormalization-group variant developed for the homopolar problem.[10] Recall that an infinitesimal renormalization-group transformation consists of excluding fast modes with $\kappa > 1 - \Delta$ and simultaneously making up blocks of $1 + 2\Delta$ monomers each. By repeating the procedure of Ref. 10, it is easy to verify that the procedure described results again in an equation of type (9)—already a very important result because it proves renormalizability of the problem—and this renormalization transformation of the β_t sequence takes the form

$$\beta_t(s(1+2\Delta)) = \sum_{n=1}^{2\Delta} \beta_{t+ns}(s) + \beta_t(s)$$

$$\times [1 + \Delta\beta_{t-s}(s) h_d(s^{-1/2}) s^{-1/2} a^{-d}].$$

(11)

Here s characterizes the current step of the renormalization and has the meaning of the number of bare (unrenormalized) monomers in one block monomer. For simplicity, we number the links in (11) in accordance with the bare monomers, regardless of the current value of s. Naturally, in this notation the renormalized monomers have numbers 0, s, $2s$,..., and the recurrence transformations of type (4) or (9) connect $G^{(t+s)}$ with $G(t)$. Owing to the critical slowing-down property, we can assume that the next block monomer starts out with the same value G_κ with which the preceding ends, since we are interested after the renormalization only in small $\kappa < 1 - \Delta$. This means that the block monomers are made up of nonoverlaping sections of the chain.

Assume now that all the β_t are statistically independent random quantities prior to the renormalization, i.e., at $s = 1$.[11] From the overlap of the neighboring block monomers it follows that the renormalized sequence β_t also consists of statistically independent quantities. Averaging (11) with this fact taken into account, we get

$$2s \frac{d\bar{\beta}(s)}{ds} = 2\bar{\beta}(s) + \bar{\beta}^2(s) h_d(s^{-1/2}) s^{-1/2} a^{-d}.$$

(12)

We see that the mean value of $\beta(s)$ is renormalized just as $\beta(s)$ for a homogeneous homopolymer.[10] The explicit solution of (12) is

$$\bar{\beta}(s) = a^d s \left[\frac{a^d}{\bar{\beta}} - \int_{s^{-1/2}}^{1} h_d(\varkappa) \varkappa^{-2} d\varkappa \right]^{-1}.$$

(13)

However, knowledge of the mean value for the heteropolymer problem is insufficient, and the question of renormalizing the probability distribution arises. We denote the distribution of the bare values of β_t by $P(\beta)$. By way of example, it is useful to consider a chain with links of two types, A and B, for which

$$P(\beta) = c_A \delta(\beta - \beta_A) + c_B \delta(\beta - \beta_B), \quad c_A + c_B = 1.$$

(14)

We denote the distribution of $\beta_t(s)$ by $P_s(\beta)$. It is convenient to write the corresponding renormalization equation in terms of the characteristic function $\tilde{P}_s(\omega)$, i.e., of a Fourier transform of $P_s(\beta)$ with respect to the variable β. To this end we must average the quantity $\exp[i\omega\beta_t(s(1 + 2\Delta))]$ with the aid of Eq. (11); this yields

$$\tilde{P}_{s(1+2\Delta)}(\omega) = \tilde{P}_s^{2\Delta}(\omega) \langle \tilde{P}_s(\omega[1 + \Delta\beta_{t-s}(s) h_d(s^{-1/2}) s^{-1/2} a^{-d}]) \rangle,$$

where the angle brackets denote averaging over $\beta_{t-s}(s)$ (recall that this quantity is statistically is independent of $\beta_t(s)$ over which we have already averaged). Separating the terms linear in Δ we obtain

$$2s \frac{\partial \tilde{P}_s(\omega)}{\partial s} = 2\tilde{P}_s(\omega) \ln \tilde{P}_s(\omega) + \frac{\partial \tilde{P}_s}{\partial \omega} \bar{\beta}(s) h_d(s^{-1/2}) s^{-1/2} a^{-d}.$$

(15)

Since $\bar{\beta}(s)$ can according to (13) to be regarded simply as a known function of s, Eq. (15) is linear in $\ln \tilde{P}_s(\omega)$ and it is easy to find its characteristics and write the explicit solution:

$$\tilde{P}_s(\omega) = \tilde{P}^s \left(\omega \exp \left[\int_1^s h_d(s'^{-1/2}) s'^{-1/2} \bar{\beta}(s') a^{-d} ds'/2 \right] \right).$$

(16)

4. ANALYSIS OF RENORMALIZATION RESULTS

The first that can be seen from (16) is a decisive change of its entire behavior when the bare value $\bar{\beta} a^{-d}$ does through the value

$$a^{-d} \beta_{tr} = \left[\int_0^1 h_d(\varkappa) \varkappa^{-2} d\varkappa \right]^{-1} = \begin{cases} (2\pi/d)^{d/2}/\zeta(d/2), & d > 2 \\ 0, & d < 2 \end{cases},$$

(17)

that corresponds to the point of the coil-globule transition in the homopolymer (the explicit expression with $d > 2$ is a Gaussian g-function). In fact, as β goes through β_{tr}, the behavior of $\beta(s)$ changes radically: The quantity $\beta(s)s^{-d/2}a^{-d}$ remains bounded for any s if $\beta \leqslant \beta_{tr}$, but diverges at finite s if $\beta > \beta_{tr}$. Obviously, the behavior of $P_s(\beta)$ changes just as drastically. This leads already to an important conclusion: the critical point of the coil-globule phase transition in an infinite statistical heteropolymer is determined by the condition

$$\bar{\beta} = \beta_{tr}.$$

(18)

To emphasize the nontrivial character of the result (18), we call attention to the following: a periodic heteropolymer with the same ratio of the components C_A and C_B goes through a coil-globule transition not under the condition (18), but at another point. Consider, for example, a periodic polymer ABAB.... (for which, of course, $C_A = C_B = \frac{1}{2}$) and assume that $\beta_A = 0$. We obtain obviously simply a homopolymer with bonds twice as long, which is described by the usual homopolymer equations with the substitution $g_k \to g_k^2$, i.e., $a \to 2^{1/2}a$, and for which the transition point is determined by the condition $\beta_B = 2^{d/2}\beta_{tr}$. In the same situation

$(C_A - C_B = \frac{1}{2}, \beta_A = 0)$ Eq. (18) gives for a statistical sequence a different answer: $\beta_B = 2\beta_{tr}$.

The foregoing example illustrates a very interesting property of the system. In fact, in a long-loop situation the heterogeneity of the sequence β_t is effectively smoothed out; this is qualitatively quite clear. But the smoothed value of β_{eff} actually obtained is determined by the small-scale properties of chain. Thus, it can be seen from (13) and (16) that the behavior of $\bar\beta(\beta)$ and $P_s(\beta)$ becomes universal only at large s, and is determined only by the asymptote of $h_d(\kappa)$ at small κ. The behavior of these quantities at small s, however, depends on the entire function $h_d(\kappa)$, i.e., on the small-scale details of the arrangement of the "bonds" $g(x)$. The small-scale properties of the inhomogeneous sequence β_t, as seen from the example with the periodic polymer, are also very important for the determination of β_{eff}. The result (18) states thus that $\beta_{eff} = \bar\beta$ for a statistical sequence.

As to the form of the $P_s(\beta)$ distribution, it is easy to conclude by analogy with the central limit theorem,[12] where the analog of (16) takes the form $\widetilde{P}_s(\omega) = \widetilde{P}_s(\omega)$, that as $s \to \infty$ the distribution becomes asymptotically normal. The corresponding mean value is given by Eq. (13), and for the variance we get from (16) the following expression, which is valid for all s and d:

$$[\overline{\beta^2}(s) - \bar\beta^2(s)]/\bar\beta^2(s) = s^{-1}(\overline{\beta^2} - \bar\beta^2)/\bar\beta^2. \tag{19}$$

This expression takes the same form as, according to the central limit theorem, the analogous equation for ordinary values of β averaged over blocks of length s. The substantial difference is, however, that $\bar\beta(s)$, and with it according to (19) also $\overline{\beta^2}(s)$, can in our situation be singular.

The $P_s(\beta)$ distribution becomes thus monomodal and its variance (19) becomes small at large s. The length of a block monomer in a coil or at the transition point can be increased all the way to the total length N of the chain. In a globule, however, there is a finite correlation radius along the chain or, equivalently, a finite characteristic length l of the loop in the fringe of the globule. Near the phase transition we have $l \gg 1$ and this is precisely why the scale l is unique and independent of the realization of the random chain or of the coordinate along the chain, because the $P_s(\beta)$ distribution is narrow at $s \sim l \gg 1$. The size of the scale l, since it is unique, is determined simply by the mean value of $\beta(s)$. A simple analysis shows that renormalizations in the globular regime preserve an accuracy of the order of unity all the way to a scale such that

$$\bar\beta(s \sim l) l^{-d/2} a^{-d} \sim 1. \tag{20}$$

The meaning of this estimate is made clear by the fact that $l^{d/2} a^d$ is the unperturbed Gaussian volume of the block monomer, and $\beta(l)$ is the virial coefficient of its interaction with the external field. Using (10) for the analysis of the asymptotic of the integral (13) at a small lower limit, we easily get

$$l(\beta) \sim \begin{cases} \left[\dfrac{\pi}{2}(d-2) a^d \left(\dfrac{1}{\beta_{tr}} - \dfrac{1}{\beta} \right) \right]^{-2/(d-2)}, & 4 > d > 2 \\[2mm] \exp(\pi a^2/\beta), & d = 2 \\[2mm] \left[\dfrac{\pi}{2}(2-d) \dfrac{a^d}{\beta} + 1 \right]^{2/(2-d)}, & d < 2. \end{cases} \tag{21}$$

The scale in a globule can thus be increased only up to $l(\bar\beta)$. Is this enough to make the variance of $\beta(s)$ small? This question is vial if $d < 2$, when the right hand side of (19) has near the transition point a large factor $\bar\beta^{-2}$ besides the small factor s^{-1}. Equation (20), however, shows that $l^{-1}\bar\beta^{-2} \sim l^{1-d}$, therefore the variance of $\beta(s \sim l)$ becomes small near the transition point at $d > 1$ but not at $d < 1$. We shall consider hereafter just the case $d > 1$.

Thus, at $d > 1$ the loops in a coil and near the point of transition to a globule turn out to be quite long, and the heterogeneity of the sequence of the links evens out over their length. The probability distribution for $\beta(s)$ becomes thus monomodal, even if the distribution $P(\beta)$ was initially bimodal as in (14), or even more complicated. This circumstance has an important physical meaning: notwithstanding the presence in the heteropolymer chain of several species of distinctly different links, we are unable to classify sufficiently long sections of the chain, in a physically meaningful manner, into qualitatively different types: the chain cannot be regarded as a sequence of globulized and coil-like segments with different fluctuations of the density of the link types.

5. CALCULATION OF THE DIMENSIONS OF A RANDOM HETEROPOLYMER

The starting point of the analysis of the spatial structure of a chain should be a recurrence relation for the ψ function (5). For example, the mean squared[2] distance from the origin O to the link t is equal to

$$R_t^2 = \int \mathbf{x}^2 \psi_t(\mathbf{x}) \, d^d x = -(\Delta_k \psi_t(\mathbf{k}))_{k=0} = -a^2 \, \partial \psi_t(\kappa)/\partial \kappa^2 \tag{22}$$

and for pointlike fields (2) we easily obtain

$$R_{t+1}^2 = (R_t^2 + a^2) \left[1 + (\beta_t/a^d) \int_0^1 h_d(\kappa) \psi_t(\kappa) \, d\kappa \right]^{-1}. \tag{23}$$

A. Coil regime. Using once relation (23) renormalized to $s = N$, and assuming that $\psi_t(\kappa)$ in this equation pertains to the start of the chain, i.e., $\psi_t(\mathbf{x}) = \psi_0(\mathbf{x}) = \delta(\mathbf{x})$, $\psi_t(\kappa) = 1, R_t = 0$, we obtain directly a characteristic of the entire chain—the mean squared distance R_N between its ends:

$$R_N^2 = Na^2 [1 + \beta_{t=0}(s = N) N^{(1-d)/2} h_d^*/da^d]^{-1}$$

$$\left(\int_0^1 h_d(\kappa N^{-1/2}) \, d\kappa \text{ for } N \gg 1 \text{ is evaluated using (14)} \right). \tag{24}$$

A. Yu. Grosberg and E. I. Shakhnovich   1287

Since $\bar{\beta}(s)$ is renormalized just as $\beta(s)$ for a homopolymer, we get directly from (24)

$$(\overline{R_N^{-2}})^{-1/2} = R_N^{(\text{homo})}(\bar{\beta}),\qquad(25)$$

where $R_N^{(\text{homo})}$ is the homopolymer dimension calculated in Ref. 10. To calculate the mean values of the other powers[3] of R_N it is necessary to expand in (24) in powers of the second term in the square brackets, and to use expression (19). For example,

$$\bar{R}_N \cong R_N^{(\text{homo})}(\bar{\beta})\left[1 + \frac{\overline{\beta^2}-\bar{\beta}^2}{\bar{\beta}^2}\frac{3h_d^{*2}}{8d^2}\bar{\beta}^2(N)N^{-d-1}a^{-2d}\right]$$

$$= R_N^{(\text{homo})}(\bar{\beta})\left[1 + \frac{\overline{\beta^2}-\bar{\beta}^2}{\bar{\beta}^2}\frac{3h_d^{*2}}{8d^2}N^{1-d}\begin{cases}\left(\dfrac{a^d}{\bar{\beta}}-\dfrac{a^d}{\beta_{tr}}+\dfrac{h_d^*}{d-2}N^{-(d-2)/2}\right)^{-2},\ d>2\\[2mm]\left(\dfrac{a^d}{\bar{\beta}}-\dfrac{h_d^*}{2}\ln N\right)^{-2},\ d=2\\[2mm]\left(\dfrac{a^d}{\bar{\beta}}-\dfrac{h_d^*}{2-d}N^{(2-d)/2}\right)^{-2},\ d<2\end{cases}\right],\qquad(26)$$

$$DR_N \equiv \left[\frac{\overline{R_N^2}-\bar{R}_N^2}{\bar{R}_N^2}\right]^{1/2}\cong\left[\frac{\overline{\beta^2}-\bar{\beta}^2}{\bar{\beta}^2}\right]^{1/2}\frac{1-[\bar{R}_N^2/Na^2]}{2N^{1/2}}\quad(\beta<\beta_{tr}).\qquad(27)$$

The second term in the square brackets of (26) is the fluctuation correction. It is, of course, the first term of a series but, most importantly, by far not in powers of the variance of β, which is not assumed small, but in powers of N^{-1}. The variance ($\sim N^{-1/2}$) of the size of the coil (27) is also small, and is additionally decreased when the coil comes close to an ideal one (for which $R_N = Na^2$). The latter is due to the smallness (at $d > 1$) of the fraction of the coil links that interacts with the local field.

At $\bar{\beta}<\beta_{tr}$ (for a coil) the results (26) and (27) are asymptotically correct at $N\gg 1$; moreover, Eq. (25) is rigorously exact. These results remain valid also at $\bar{\beta}>\beta_{tr}$ for relatively short chains $N\ll l(\bar{\beta})$ (or for chains with N given in the immediate vicinity of the transition point).

B. Globular regime. In a truly globular situation we should have not only $\bar{\beta}>\beta_{tr}$ but also $N\gg l(\bar{\beta})$. The divergence of $\bar{\beta}(s)$ (as well as of $\beta^2(s)...$) as $s\sim l(\bar{\beta})$ becomes therefore essential. Just as in a homopolymer, the scale $l(\bar{\beta})$ is none other than the characteristic loop length of the globule. The inequality $N\gg l(\bar{\beta})$ means that globulized chain contains many loops, and the divergence is interpreted as a manifestation of the fact that at scales $s>l(\bar{\beta})$ the chain is strictly held together by the external field. For a heteropolymer, however, one can go even farther: it is natural to assume that over the scale on which the block links are rigorously held together in the globule, the probability distribution $\mathscr{P}_t\{\psi\}$ (8) is stabilized over a length of the order of one block link. To explain this fact, let us examine relation (9) once more and assume that the renormalization in it has already been advanced to a relatively large scale $s\gg 1$. If at the same time the function $G_t(x)$ is localized over distances $\sim as^{1/2}$, then $G_{t+s}(x)$ becomes practically independent of the form of $G_t(x)$ and is determined only by the integral Z_t and by the primary structure of the chain on the section from t to $t+s$. Accordingly, $\psi_t(x)$ should become independent of $\psi_{t+s}(x)$ at a value of s such that $as^{1/2}$ becomes comparable in order of magnitude with the size of the localization region $\psi(x)$, the latter being $\sim al^{1/2}$ in a globule.

We arrive thus at the conclusion that the mutual influence of the functions $\psi_t(x)$ does not extend farther than a length l along the chain. In particular, the value of R_N for a globule is independent of N and is determined only by the last section of a chain consisting of $\sim l$ links. To calculate this value we must use Eq. (23), renormalized to $s\sim l$, to obtain R_{t+l}, regardless of what is used here for ψ_t; (it is simplest to put $\psi_t(x) = \delta(x)$). As a result we obtain for R_N a formula of the same form as (24), but containing l in place of N and $\beta_{t=N-l}(s\sim l)$ in place of $\beta_{t=0}(s=N)$:

$$R_N^2\approx a^2 l[1+\beta_{t=N-l}(s\sim l)l^{-d/2}h_d^*/da^d]^{-1}.\qquad(28)$$

In contrast to Eq. (24) which is exact for the coil, its analog (28) for the globule is accurate to unity. Continuing the calculations at this accuracy, it is natural to expand in (28) in powers of the second term in the square bracket; terms of higher order in β^2 turn out to be small here by virtue of $l\gg 1$. As a result we obtain the mean value and the variance of R_N for a set of globulized random chains:

$$\bar{R}=R^{(\text{romo})}(\bar{\beta})\left[1+\text{const}\frac{\overline{\beta^2}-\bar{\beta}^2}{\bar{\beta}^2}l^{-1}(\bar{\beta})\right],\quad R^{(\text{romo})}\sim al^{1/2}(\bar{\beta}),\qquad(29)$$

$$DR\equiv\left(\frac{\overline{R^2}-\bar{R}^2}{\bar{R}^2}\right)^{1/2}\sim l^{-1/2}(\bar{\beta})\left(\frac{\overline{\beta^2}-\bar{\beta}^2}{\bar{\beta}^2}\right)^{1/2}\qquad(30)$$

under the condition $N\gg l(\bar{\beta})$. We omit for $\bar{R}$ and $\overline{R^2}$ the subscript N, on which they do not depend in the globular regime. We have used the equal sign in (29) in view of the continuity of the dependence of R on the variance of β; the fluctuation term in $\bar{R}$, just as the variance DR, were determined by us only apart from a factor. Just as in (26), the expansion in (29) is not with respect to the variance of β, which can be arbitrary. The small parameter in (29) is in fact $l^{-1}(\bar{\beta})$, i.e., the proximity to the transition point.

We see that the mean distance $\bar{R}_N$ between the ends of a heteropolymer chain practically coincides in the globular regime (29), just as in the coil regime (26), with the analogous distance for an "average" homopolymer with $\beta = \bar{\beta}$. The fluctuational correction term in (29) for $d > 1$ is small near the transition point if $l(\bar{\beta}) \gg 1$, as is the analogous term in (26) for a coil. It must be understood, however, that on moving farther from the transition point into the globular region the dependence of R on the primary structure becomes stronger, and the difference between $\bar{R}$ and $R^{(homo)}$ not only increases qualitatively but becomes also qualitative.

The simplest characteristic of the sensitivity of the spatial structure of a chain to a change of a sequence of types of links is the variance of the dimension. Note that in the coil regime DR tends to zero as $N \to \infty$ (27). On the contrary, in a globule DR tends as $N \to \infty$ to a finite limit. On the whole, the value of $\lim DR$ as a function of $\bar{\beta}$ has a singularity at $\bar{\beta} = \beta_{tr}$, i.e., at the end of the transition. The character of the singularity is determined by Eqs. (21) and (30).

We call particular attention to the case $2 < d < 4$, which is most important in practice. In this case the dependence of DR on $\bar{\beta}$ has in the globular region a maximum at

$$\bar{\beta} = \beta_{tr}(d-1)/(d-2), \tag{31}$$

i.e., at a finite distance from the transition point. A similar dependence was obtained earlier in a computer experiment.[11] Thus, the theoretical result shown in Fig. 1 agrees with the results of the computer experiment.

The quantity R characterizes by definition precisely the end point of the chain. In a globule, however, the loop sizes are also of the order of R, and moreover, the probability distributions for them is the same as for R. Therefore R (29) is the average dimension of a loop in one long globulized chain, and DR (30) is the relative scatter of the dimensions of various loops in one chain. The number of loops in a chain is $\sim N/l(\beta)$, since the fraction of particles in the core of the globule is small near the transition. The radius of inertia of the globule,[4] which characterizes it as a unit, is a self-averaging quantity: it receives statistically independent contributions from a large number of loops. Accordingly,

$$\bar{\mathscr{R}} \sim \bar{R}, \quad D\mathscr{R} \sim DR(N/l(\beta))^{-1/2}. \tag{32}$$

6. CALCULATIONS OF THE FREE ENERGY OF A RANDOM HETEROPOLYMER

The general expression for the free energy (7) for pointlike external fields (2) is reducible, according to (23), to the form

$$F_N = -T\sum_t \ln[(R_t^2 + a^2)/R_{t+1}^2] = -T\sum_t \ln(1 + a^2/R_t^2), \tag{33}$$

where we have left out the term $\ln(R_1/R_N)^2$, which is insignificant in the thermodynamic limit (it is $\lesssim \ln N$).

We determine first the mean value of the free energy of a random globule. To this end it is necessary in (33) to expand in powers of (a^2/R_t^2) and then average each term of the expansion in the manner described in the derivation of (2). Since the R_t are uniformly distributed, according to (28), in a globule for almost all the links (except at the start of the first loop), we easily obtain

$$\bar{F}_N = F_N^{(homo)}(\bar{\beta})\left[1 + \text{const}\frac{\overline{\beta^2} - \bar{\beta}^2}{\bar{\beta}^2}l^{-1}(\beta)\right],$$

$$F_N^{(homo)}(\bar{\beta}) \sim -TN/l(\beta). \tag{34}$$

A clarification is in order here. If (33) is formally averaged with the aid of (28), the fluctuation term contains l^{-2} in place of the correct l^{-1}, because (a^2/R_t^2), the main term in (33), is linear in $\beta(s = l)$ in the approximation (28). In fact, however, the absence of a term quadratic in $\beta(s = l)$ in (28) is due not to some fundamental cause [as in (24)], but to the fact that (28) is approximate. We find therefore that the fluctuation term in $\overline{R^{-2}}$, i.e., in the free energy, is of the same order for the globule as in $\overline{R^\alpha}$ with some other α, as indeed written in (34). Thus, the fluctuation correction to the free energy of the globule (34), just as to its size (29), becomes small as the transition point is approached for any $d > 1$. The analogous fluctuation terms for a coil turn out to be proportional to negative powers of N. Such terms are meaningless in the free energy, i.e., in a coil we have, as also for a homopolymer, $F_N = 0$.

The result (34) has an important physical meaning. It means that not only the transition point (18), but also the thermodynamic type of transition in an infinite heteropolymer is the same as in an "average" homopolymer. In other words, introduction of an inhomogeneity in a linear structure of a polymer leaves the coil-globule transition in the same universality class to which it belongs for a homopolymer (at $d > 1$).

The probability distribution for the free energy of a globule is normal, since the chain contains many loops, $N \gg l(\bar{\beta})$. In fact, the sequence of terms in (33) can be broken up in blocks containing l terms each, and these blocks
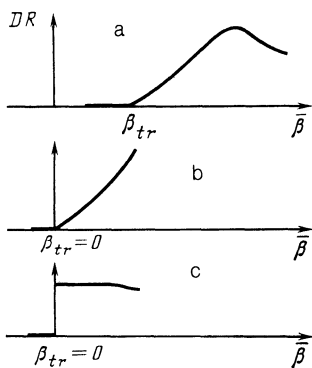

FIG. 1. Dependence of the variance of the inertia radii of random chains on $\bar{\beta}$ for $\overline{\beta^2} - \bar{\beta}^2 = \text{const}$ and for different dimensionalities: a) $2 < d < 4$, b) $1 < d < 2$, c) $d \to 1$.

are practically independent of one another. To estimate the variance of the free energy we must reason as follows. The terms within the blocks are strongly correlated and the spread of their sum is of the same order as the relative spread of one term a^2/R_t^2, and this quantity is easily calculated with the aid of (28) and (19). The addition of the contributions of the blocks, however, is subject to the central limit theorem. Simple calculations yield the following very simple result:

$$\frac{\overline{F_N^2} - \overline{F_N}^2}{\overline{F_N}^2} \sim \frac{1}{N} \frac{\overline{\beta^2} - \overline{\beta}^2}{\overline{\beta}^2}, \qquad (35)$$

which is independent of the dimensionality of the space.

Equation (35) allows us to find the probability of losing the thermodynamic stability, defined by the inequality $F_N < 0$, of the globular state at $\beta \neq \beta_{tr}$. Since F_N has a Gaussian distribution, the probability in question is small under the condition

$$N \gg \overline{\beta}^2/(\overline{\beta^2} - \overline{\beta}^2). \qquad (36)$$

Consider an example in which $\overline{\beta^2} - \overline{\beta}^2$ remains constant when $\overline{\beta}$ is varied. At $d > 2$, for example in the most important three-dimensional case, the condition (36) reduces in the transition region simply to $N \ll 1$, because $\overline{\beta} > \beta_{tr} > 0$. At $d < 2$, however, when $\beta_{tr} = 0$, it turns out that the free energies of different heteropolymers of length N (differing in the sequence of the elements) vanishes in the interval $|\overline{\beta}| \lesssim N^{-1/2}$. In a finite chain, however, it is meaningless to speak of two different states at $|F_N| \lesssim T$, i.e., the transition becomes smeared out in an interval of width $|\overline{\beta}| \lesssim N^{-(2-d)/2}$. A comparison shows that at $d > 1$ the usual thermodynamic smearing of the transition point exceeds the smearing due to the uncertainty of the primary structure.

7. CONCLUSION

We have thus obtained a sufficiently complete solution of the problem of collapse of an immaterial heteropolymer under the influence of a pointlike external field. Recall that the results (25)–(27) for the coil phase are exact. Moreover, they are asymptotically exact in the limit $N \to \infty$ also for a nonpointlike external field. Also exact are the results that the transition point (18) and the transition order (34) are the same for a stochastic heteropolymer and an "average" homopolymer. The results (29), (30), (34), and (35) for a globule near a phase transition are valid accurate to numerical factors.

We have not considered the regime of a well shaped globule with short loops far from the transition point. The point-field model is obviously senseless for the description of this regime, and the renormalization-group method is useless. Yet a highly varied behavior of a globule and interesting peculiarities can be expected in this region. This is clear, for example, from a comparison with the data on helix-coil transitions,[13] where it is precisely for moderate loop lengths that one observes the most interesting effect, viz., successive "melting out" of individual segments of the chain. Of course, the analogy with the helix-coil transition is complete only for point fields, and for physically natural extended field the behavior of the globule should be even more varied. Although there are three exactly solvable models of heteropolymer globules for definite field shapes,[8,14,15] the question of their description far from the transition into a coil remains on the whole open and timely.

Even more important and quite subtle is the question of the qualitative conclusions that can be drawn from the results of the present paper for proteins. We discuss this in part in Ref. 11.

---

[1]We do not consider here the subtle question of the applicability of a random primary structure to simulation of biopolymers. For details and a bibliography on this subject see Ref. 11.

[2]We have in mind, of course, a thermodynamic average over all states of the given random chain.

[3]To prevent misunderstandings, we point out that in our notation $R_N^\alpha \neq \int |x|^\alpha \psi_N(x) d^d x$ for any $\alpha \neq 2$, and $R_N^\alpha = (R_N^2)^{\alpha/2}$, where R_N^2 is defined in (22).

[4]The square of the inertia radius is the ratio of the moment of inertia to the mass.

---

[1]I. M. Lifshitz, Sov. Phys. JETP **55**, 2408 (1968) [Sov. Phys. JETP **28**, 1280 (1969)].

[2]I. M. Lifshitz, A. Yu. Grosberg, and A. R. Khokhlov, Rev. Mod. Phys. **50**, 683 (1978).

[3]I. M. Lifshitz, A. Yu. Grosberg, and A. R. Khokhlov, Usp. Fiz. Nauk **127**, 353 (1979) [Sov. Phys. Usp. **22**, 123 (1979)].

[4]T. M. Birshteĭn, Macromolecules **12**, 715 (1979).

[5]A. A. Vedenov, A. M. Dykhne, and M. D. Frank-Kamenetskiĭ, Usp. Fiz. Nauk **105**, 479 (1971) [Sov. Phys. Usp. **14**, 715 (1972)].

[6]A. Yu. Grosberg, Preprint, Preprint, Scient. Center for Biol. Research, Pushchino-na-Oke, 1983.

[7]S. P. Obukhov, Nordita Preprint, 1984.

[8]A. Yu. Grosberg, J. Stat. Phys. **38**, 149 (1985).

[9]G. M. Zaslavskiĭ, Stochasticity of Dynamic Systems [in Russian], Nauka, 1984.

[10]A. Yu. Grosberg and E. I. Shakhnovich, Zh. Eksp. Teor. Fiz. **91**, 837 (1986) [Sov. Phys. JETP **64**, 493 (1986)].

[11]A. Yu. Grosberg and E. I. Shakhnovich, Biofizika **31**, No. 6 (1986).

[12]V. Feller, Probability Theory and Its Applications [Russ. transl.], Mir, 1983.

[13]Yu. S. Lazurkin, Molekulyar. Biolog. **11**, 1311 (1977).

[14]A. Yu. Grosberg, E. M. Kaganova, and S. A. Molchanov, Biofizika **29**, 40 (1984).

[15]A. Yu. Grosberg and E. M. Kaganova, Dokl. Akad. Nauk SSSR **290**, No. 4 (1986).

Translated by J. G. Adashko