State diagram of a polymer globule and the problem of self-organization of its spatial structure

I. M. Lifshitz and A. Yu. Grosberg

Institute of Physics Problems, USSR Academy of Sciences;
Research Institute for Biological Tests of Chemical Compounds, Ministry of Medical Industry

(Submitted June 26, 1973)
Zh. Eksp. Teor. Fiz. 65, 2399–2420 (December 1973)

It is shown by means of statistical thermodynamics that a globule made up of a long polymer chain, depending on the length of the chain and on the temperature, can be either surrounded by a loose gaslike shell, or can have a sharp boundary outside of which the monomer density vanishes. Transitions between these states and also transitions from these states to the state of a loose random coil are first-order phase transitions. A detailed analysis is carried out of the phase states and of the transitions between them for the model of "beads interacting on a flexible filament." The corresponding phase diagram is constructed. The question of the reliability of the self-organization of a definite tertiary structure of a biopolymer globule is considered both from the thermodynamic and from the kinetic points of view. It is shown that from among all the possible sequences of monomers along the chain there is an exceedingly small fraction of those capable of forming a tertiary structure similar in its ordering to an aperiodic crystal, and capable by the same token of functioning biochemically.

A long biopolymer chain is, by itself, a macroscopic system. This makes it possible to describe it by statistical-physics methods and to consider the organization of its spatial tertiary structure as a phase transition from the coil state to the globular state.

A consistent statistical-thermodynamic theory of polymer chains was developed by I. M. Lifshitz.[1] By way of example of application of the theory, the coil-globule phase transition was considered for an extremely long chain. The present paper is devoted to a detailed study of the structure of different globules (both large, in which the role of the surface is relatively small, and small, in which the "surface" and "volume" effects are of the same order of magnitude) and of the corresponding phase transitions.

The microscopic theory of real biological globules is highly unique and is determined by the primary structure of the chain. Of course, the statistical description is not aimed at determining such details of the structure of a particular globule. Nonetheless, many qualitative properties of real globules become clear even from a simple model theory, and we shall see, in addition, that this approach enables us to determine the conditions needed for the reliability of the self-organization of a strictly ordered tertiary structure, and by the same token for the possible occurrence of a native state of the globule.

The present paper is entirely based on the results of[1] and is in fact its continuation. We therefore begin with a brief qualitative exposition of the principal premises of[1] in a form that is most convenient for the sequel.

1. DESCRIPTION OF BASIC MODEL

The macroscopic state of a polymer chain suspended in a solution (at fixed initial conditions) is a partial-equilibrium state, in which the primary structure, i.e., the sequence of links along the chain, is assumed to be fixed and to form the "linear memory" of the system (see the Appendix). This memory can be described by correlations g_j between neighboring links. In the general case, g_j depends on the coordinates, orientations, and the internal properties of links numbered j and j + 1. We confine ourselves, however, to a very simple model in which the linear memory involves only memorizing of the location of each link in the linear sequence along the chain, and all the correlations g_j reduce to a single function $g(|x_j - x_{j+1}|)$, which depends only on the distance between the neighboring links.1) An illustrative sample of such a system is an elastic flexible filament on which beads are strung; this picture must always be kept in mind in the course of reading this paper (Fig. 1).

Consideration of this model is expedient for two reasons. First, such a simple model theory clarifies many qualitative features of the more general case. Second, the model in question may be directly applicable to a heteropolymer molecule, in which certain monomers interact much more strongly than the others, for example to a flexible polymer chain with relatively widely spaced but strongly interacting "appendages." In this situation, the role of the "link" should be played by the entire segment of the chain between two strongly interacting monomers.2)

2. FUNDAMENTAL PREMISES OF THE STATISTICAL-THERMODYNAMIC THEORY OF POLYMER CHAINS [1]

The collapse of a chain to form a dense globule can be due to the action of an external compressing field and (or) to forces of "lateral" ("volume") interaction. The role of the external field can be played by a local change in the composition or in the state of the solvent. By volume interaction we mean lateral hydrogen bonds, Van der Waals, multipole, polarization, and other weaker interactions of links that are far from each other along the chain, but come close together as a result of bending of the chain.

We denote by $\varphi(x)$ the external field and by $\epsilon(\Gamma)$ the

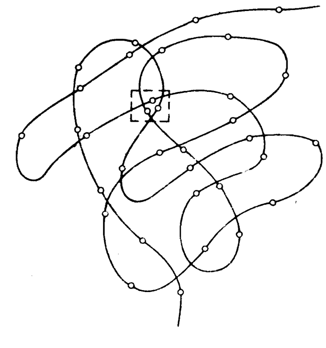

FIG. 1. "Interacting beads on a flexible filament" model. The volume element $v \ll a^3$ is framed; the particles inside this volume interact with one another but are very far from one another along the chain.

energy of the volume interaction ($\Gamma = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ is a point in the configuration space of the system). We emphasize once more that $\epsilon(\Gamma)$ does not include the energy of the strong longitudinal bonds, which are taken into account by the correlations g_j. The statistical equilibrium of a system with linear memory is described by the distribution

$$\rho(\Gamma)\Delta\Gamma=\exp\left\{\frac{1}{T}\left[F-\varepsilon(\Gamma)-\sum_{j=1}^{N}\varphi(x_j)\right]\right\}\prod_j g_j\Delta\Gamma \qquad (2.1)$$

(F is the free energy of the chain$^{3)}$), which differs from the ordinary Gibbs distribution in the presence of the factor Πg_j.

We consider now a volume element $v \ll a^3$, where a is the distance between links (beads) along the chain, i.e., in other words, the characteristic length over which the correlation function $g(\mathbf{x})$ decreases:

$$a^2=\frac{1}{6}\int y^2 g(\mathbf{y})\,d^3y.$$

The longitudinal bonds (the thread) impose practically no limitations on the possible displacements of the particles (beads) within the limits of this element. On the other hand, if the number of particles in the considered volume is large enough, $nv \gg 1$ (n is the particle-number density), then it is clear that the picture of their lateral interactions is the same as if they were not bound into a chain at all. It is therefore natural to specify the macroscopic state of the chain in terms of a density distribution $n(\mathbf{x})$ that has been "smoothed out" over scales smaller than a and varies over distances $\sim$a (see the Appendix). Then the foregoing statement can be formulated in the following manner: in a macroscopic state specified by a density distribution, the free-energy change due to interaction with the external field coincides with the change occurring in a system in which the same particles are not bound into a chain.

For an arbitrary form of interaction, even in an ordinary system of particles without linear memory, there is no general method of calculating the free energy in the condensed phase. However, the arguments advanced above lead to the following natural formulation of the problem. We assume that we know the arbitrary thermodynamic characteristics of an equivalent system of links having the same volume interaction $\epsilon(\Gamma)$ but disconnected bonds (i.e., without a linear memory), and attempt to construct the statistical thermodynamics of a polymer chain in terms of these characteristics. To be specific, let f(n, T) be the free energy per unit volume of the equivalent system of disconnected links with density n and temperature T, $\mu(n, T) = \partial f/\partial n$ of the chemical potential of this system, and $p(n, T) = n\mu - f$ its pressure. We introduce the quantities

$$f^*=f-f_{id},\quad \mu^*=\mu-\mu_{id},\quad p^*=p-p_{id},$$

where f_{id}, μ_{id}, p_{id} are respectively the free energy per unit volume, the chemical potential, and the pressure of an ideal gas of links ($\mu_{id} = T \ln n$).

In this notation, the free-energy change due to interaction and to the external field is

$$E\{n(\mathbf{x})\}=\int\{f^*(n(\mathbf{x}))+\varphi(\mathbf{x})n(\mathbf{x})\}\,d^3x. \qquad (2.2)$$

This quantity does not coincide, of course, with the system energy, but will play the role of energy in the theory that follows (see the Appendix). The result (2.2), as shown in[1], holds in both limiting cases, both when the radius r_0 of the lateral interaction is small ($r_0 \ll a$), so that almost all the particles that interact with one another belong to remote sections of the chain, and when the interaction radius is large ($r_0 \gg a$), so that any particle interacts with a practically constant number of other particles. We can therefore assume that the result (2.2) is valid, at least qualitatively, also for all the intermediate cases, i.e., for any real interaction.

To express the configuration entropy of the system explicitly as a functional of a specified nonequilibrium density distribution, we proceed in the following manner. We consider a chain without volume interaction and place it in an external field in which the specified density distribution turns out to be the equilibrium distribution. For such a chain we can obtain the free energy (by directly calculating the statistical integral of the canonical distribution, see[1]). Since the energy (2.2) is known, we obtain by the same token the configuration entropy (see the Appendix). As a result we get

$$S\{n(\mathbf{x})\}=\int n(\mathbf{x})\ln\frac{\hat{g}\psi}{\psi}\,d^3x, \qquad (2.3)$$

where the integral operator $\hat{g}$ is determined by the equation

$$\hat{g}\psi=\int g(\mathbf{x}-\mathbf{x}')\psi(\mathbf{x}')\,d^3x', \qquad (2.4)$$

and the function $\psi(\mathbf{x})$ coincides, apart from a normalization factor, with the distribution function of the terminal link of the chain and is expressed in terms of $n(\mathbf{x})$ in the form

$$\Lambda n(\mathbf{x})=\psi(\mathbf{x})\hat{g}\psi. \qquad (2.5)$$

The constant Λ determines the normalization of the ψ function (the choice $\Lambda = 1$ was made in[1]). The meaning of these results becomes clear if it is noted that $n(\mathbf{x})v$ particles can be situated in a volume v near the point **x** only if $\sim$nv particles are distributed in one manner or another about the point x within a distance $\sim$a; the expression (2.5) for the density therefore contains the operator $\hat{g}$. The same circumstance can also be taken into consideration when calculating the number of microscopic states realizing the specified distribution of the smoothed-out density (i.e., in essence, the configuration entropy). More briefly speaking, the nonlocality of the entropy (2.3) due to the operator g is a direct manifestation of the linear memory. It is precisely this nonlocality which leads actually to the principal singularities in the statistics of polymer chains.

To find the equilibrium density distribution, we must minimize the free energy (see the Appendix)$^{4)}$

$$F\{n\}=E\{n\}-TS\{n\}$$

under the additional condition that the total number of particles be constant

$$N=\int n(\mathbf{x})\,d^3x. \qquad (2.6)$$

As a result we get the equation

$$\hat{g}\psi=\psi\exp\left\{\frac{1}{T}\left[\mu^*(n,T)+\varphi(\mathbf{x})-\lambda\right]\right\} \qquad (2.7)$$

(λ is a Lagrangian multiplier), which must be solved simultaneously with (2.5). The solution of (2.7) must satisfy the natural requirement that its sign be constant. We see that the approximation (2.2) chosen for the energy is equivalent to the use of the quantity $\mu^*(n(\mathbf{x}))$ as the effective self-consistent field.

Equations (2.5)–(2.7) constitute a complete system for the determination of the equilibrium density distribution and of the constant λ.

For the equilibrium value of the free energy, which is determined by substituting (2.7) in (2.2)−^{9}2.3), we have

$$F=N\lambda-\int p^{*}(n,T)\,d^{3}x. \qquad (2.8)$$

3. THREE MACROSCOPIC PHASES OF A POLYMER CHAIN

At a relatively high temperature (or in the absence of interaction and of external fields), the state of the polymer chain is that of a loose statistical coil. The characteristic property of the coil state consists in the fact that while the coil does have an average density, the density fluctuations are of the order of the density itself. In other words, the coil executes continuous macroscopic pulsations, and the density is not a thermodynamically reliable quality.

When an external compressing field or a sufficiently large volume interaction is turned on, the system goes over into a "condensed" state, namely a globule, the structure of which is determined by the particular character of the interaction or of the external fields. In the globular state, the density is thermodynamically reliable and the results (2.2)−(2.8) therefore become applicable.

In view of the isotropy of the correlation function, i.e., in essence, in view of the assumed flexibility of the filament, it is natural to expect the globule produced in the absence of external fields to consist of a dense "condensed" core of spherical shape, around which the density gradually decreases to zero (Fig. 2a).

Normalizing the ψ function by means of the choice

$$\Lambda=\exp\{-\lambda/T\},$$

we can rewrite (2.5) and (2.7) in the form

$$\psi=n\exp\left\{-\frac{\mu(n)+\varphi(x)}{2T}\right\}, \qquad (3.1)$$

$$\hat{g}\psi=\Lambda\exp\left\{\frac{\mu(n)+\varphi(x)}{2T}\right\}. \qquad (3.2)$$

It is seen directly from (3.2), in particular, that the total local chemical $\mu+\varphi$ should be a continuous function of the coordinates, since the function $\hat{g}\psi$ is continuous (and is even smooth to the extent that the kernel of the operator is smooth), regardless of the continuity and smoothness of $\psi(\mathbf{x})$. The qualitative reason for this is that the particle motion within the confines of the element $v \ll a^3$ is not limited by longitudinal bonds, so that the equilibrium condition within this element reduces to the standard requirement that the local chemical potential be constant. Consequently, the local chemical potential can vary significantly only over distances $\sim a$; in particular, it can have no discontinuities in the absence of external fields.

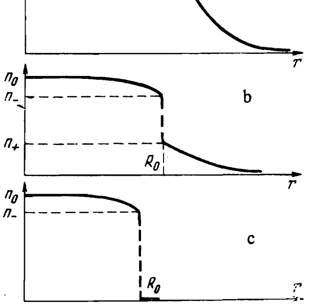

FIG. 2. a) Density distribution along the radius in a globule with fringe at a temperature exceeding the critical temperature of the system of disconnected links T > T$_c$; b) same, in the globule with a fringe at T < T$_c$; c) same, in the globule without a fringe.

However, at temperatures T lower than the critical temperature T_c of the system of disconnected links, the function $\mu(n)$ breaks up into two branches separated by a lability section. Consequently at $T < T_c$ the core of a globule produced in the absence of external fields has a sharp boundary, at which the density experiences a discontinuity (Fig. 2b), but such that the boundary values of the density n_- (on the interior side) and n_+ (on the exterior) are connected by the relation

$$\mu(n_-)=\mu(n_+)=\tilde{\mu}. \qquad (3.3)$$

This relation, which expresses the requirement of equilibrium relative to local displacements of the particles through the boundary, assumes, as is well known, the same form in the usual system without linear memory. We emphasize in this connection that the second boundary condition, which is due to the requirement that the boundary itself be in mechanical equilibrium, does not reduce in our case to a continuity of the local pressure (as stated erroneously in[1]), since the longitudinal bonds can transmit a mechanical force over a distance $\sim a$ from the boundary. In the mathematical formalism, the second boundary condition is obtained by additional minimization of the free energy (2.8) (see (3.5) below).

We note that by virtue of (2.5) the function $\psi(\mathbf{x})$ has discontinuities in the same places as $n(\mathbf{x})$. It is therefore convenient to introduce the following notation. We denote the ψ function inside and outside the core by ψ_{in} and ψ_{out}, respectively, and introduce the operators

$$\hat{g}_{in}\psi=\int_{V_0} g(\mathbf{x}-\mathbf{x}')\psi(\mathbf{x}')\,d^3x',$$

$$\hat{g}_{out}\psi=\int_{\text{outside } v_0} g(\mathbf{x}-\mathbf{x}')\psi(\mathbf{x}')\,d^3x',$$

where v_0 is the volume occupied by the core of the globule. In terms of this notation, we can write

$$\hat{g}\psi=\begin{cases}\hat{g}\psi_{in}+\hat{g}_{out}(\psi_{out}-\psi_{in}), & \mathbf{x}\text{ inside } V_0, \quad \text{(I)}\\ \hat{g}\psi_{out}+\hat{g}_{in}(\psi_{in}-\psi_{out}), & \mathbf{x}\text{ outside } V_0. \quad \text{(II)}\end{cases} \qquad (3.4)$$

The form (3.4) implies that the functions ψ_{in} and ψ_{out} lie respectively in (I) and (II). It is obvious when this form is used that Eq. (3.2) has a certain solution at any value $\tilde{\mu}$ of the chemical potential (3.3) on the boundary. It is therefore necessary to solve (3.2) with $\tilde{\mu}$ assumed as a specified parameter, and then obtain the actually realized value of $\tilde{\mu}$ by additional minimization of the free energy (2.8):

$$\partial F/\partial\tilde{\mu}=0. \qquad (3.5)$$

The concrete realization of this program turns out to be essentially different in the opposite limiting cases of large and small globules.

Extremely large globules, at temperatures above the critical temperature of the system of disconnected links ($T > T_c$) were considered in.[1] We consider now the case $T < T_c$, and assume in addition $n_- \gg n_+$, so that the density in the core of the globule is large and expression (2.2) is valid, while in the gaslike fringe the density is so small that the volume interaction of the particles in the fringe can be neglected completely.

The density in the gaslike fringe decreases with decreasing temperature, in analogy with the density of vapor over a liquid drop. The very existence of the vapor atmosphere over the liquid drop at arbitrarily low temperature is due to the fact that the vapor particles are capable of moving away to large distances (limited only by the vessel dimensions) from the liquid. In the case of a polymer chain, however, the longitudinal bonds do not permit the density of the gaslike fringe to become too small. Therefore, at sufficiently low temperature, the chain forms a globule without any gaslike fringe at all (Fig. 2c). A shorter chain can form a globule without a fringe directly from the coil.

Thus, in the absence of external fields the polymer chain can be in three different phase states: coil, globule with fringe, globule without fringe. It is our task to plot the corresponding phase diagrams in terms of the variables N and T.

4. STRUCTURE OF EXTREMELY LARGE GLOBULE WITH FRINGE

A globule made from an extremely long polymer chain (without an external field) contains a spherical core of volume $V_0 = 4\pi R_0^3/3 \gg a^3$. In the interior of the core, the density varies very slowly (assume that the corresponding constant value is n_0), and therefore $\hat{g}\psi \approx \psi$, i.e.,

$$\lambda=\mu^*(n_0, T). \tag{4.1}$$

For the free energy we therefore have

$$F=\{N\mu^*(n_0,T)-V_0p^*(n_0,T)\}+\left\{\int\limits_{V_0}p^*(n_0,T)\,d^3x-\int p^*(n,T)\,d^3x\right\}.$$

The second term is obviously proportional to the surface area of the globule of volume V_0. Therefore the density n_0 inside the core of a large globule is determined by the minimum of the first term.

Recognizing that $V_0 \approx N/n_0$, we obtain

$$p^*(n_0, T)=0 \tag{4.2}$$

(i.e., $p(n_0, T) = n_0T$). The free energy is therefore

$$F=N\mu^*(n_0, T)+F_\sigma, \quad F_\sigma=4\pi R_0{}^2\alpha, \tag{4.3}$$

where α is the surface-tension coefficient.

Proceeding to an analysis of the structure of the surface layer of a large globule, we note that the problem of the density distribution in this layer is in essence a one-dimensional problem. Introducing along the radius the coordinate $x = r-R_0$, we can, in particular, express the surface-tension coefficient in the form

$$\alpha=-\int\limits_{-\infty}^{\infty}p^*(n,T)\,dx+\alpha_0(T), \tag{4.4}$$

where $\alpha_0(T)$ is the surface tension in the system of disconnected links.

If (4.2) is satisfied, then the fundamental equation (2.6) is the condition that the surface term of the free energy be minimal. Therefore the additional minimization of (3.5) reduces to imposition of the boundary condition (4.2) as $x \to -\infty$ (in other words, we have minimized F with respect to n_0, and must now find the connection between n_0 and $\tilde{\mu}$, see (4.14) below).

To abbreviate the formulas that follow, we introduce the notation

$$\nu(\psi)=\exp\{\mu(n,T)/2T\}=n/\psi. \tag{4.5}$$

At the chosen normalization of the ψ function, ν and ψ are simply the thermodynamic variables of the system of disconnected links. They are defined by Eqs. (4.5), and it is more convenient to use them in place of μ and n.

If the function $\psi(x)$ varies smoothly enough, then the integral operator $\hat{g}$ can be approximately replaced by a differential operator, so that Eq. (3.2) reduces to the differential equation

$$a^2d^2\psi/dx^2+\psi-\Lambda\nu=0. \tag{4.6}$$

To obtain this result it is necessary in the expression (2.4) for $\hat{g}\psi$ first that ψ be taken to be independent of the coordinates perpendicular to x and that integration be carried out with respect to them and, second, that $\psi(x')$ be expanded in powers of $x-x'$ with only the first three terms retained.

At $T > T_c$, when the ψ function is continuous in the entire transition region, integration of (4.6) yields a complete solution of the problem (see[1]).

At $T < T_c$, the approximation of the integral operator by a differential one should be made more accurate near the discontinuity point of the function $\psi(x)$. It is convenient to use the notation (3.4) for this purpose. We note that the expression $g_{\text{out}}(\psi_{\text{out}}-\psi_{\text{in}})$, if taken at a point inside V_0, receives its main contribution from integration over the region near the discontinuity point, and in this region it is possible to expand the function $\psi_{\text{out}}-\psi_{\text{in}}$ in powers of x. We can proceed similarly with the expression $g_{\text{in}}(\psi_{\text{in}}-\psi_{\text{out}})$ outside V_0. As a result we find that the influence of the discontinuity reduces to the appearance, in the right-hand side of (4.6), of "surface terms" that attenuate rapidly when a distance ~a from the discontinuity is reached:

$$a^2d^2\psi/dx^2+\psi-\Lambda\nu=[\psi(-0)-\psi(+0)]\alpha_0(|x|). \tag{4.7}$$

Here and throughout we put

$$\alpha_k(x)=\frac{1}{k!a^k}\int\limits_{-\infty}^{-x}(y+x)^kg(y)\,dy. \tag{4.8}$$

Far from the discontinuity, where these surface terms vanish rapidly, Eq. (4.7) goes over into (4.6) and can therefore be readily integrated

$$\frac{x}{a}=\begin{cases}X_+(\psi)+\text{const}, & x\geqslant a,\\ X_-(\psi)+\text{const}, & x\leqslant -a,\end{cases}$$

$$X_\pm(\psi)=-\int\limits_{\bar\psi}^{\psi}\left[2\int\limits_{\psi(\pm\infty)}^{\psi}\{\Lambda\nu_\pm(\psi)-\psi\}\,d\psi\right]^{-1/2}d\psi, \tag{4.9}$$

where

$$\bar\psi=[\psi(+0)+\psi(-0)]/2, \quad \psi(+\infty)=0, \quad \psi(-\infty)=\psi_0.$$

On the other hand, near the discontinuity, where the right-hand side of (4.7) is appreciable, Eq. (4.7) can be linearized (by simply linearizing the function $\nu_\pm(\psi)$ separately on the two sides of the discontinuity). Therefore the influence of the boundary terms reduces to the appearance in $\psi(x)$ of an additive increment that can easily be calculated in explicit form. The calculations yield ultimately

$$\frac{x}{a}=\begin{cases}X_+(\chi_+), & x>0\\ X_-(\chi_-), & x<0\end{cases}; \tag{4.10}$$

$$\chi_\pm=\psi\pm[\psi(-0)-\psi(+0)]\alpha_2(\pm x),$$

where $\alpha_2(x)$ is defined by (4.8). It is assumed in (4.9) that the linear approxmations of the function $\nu_\pm(\psi)$ are continued beyond the boundary $\psi(\pm 0)$.

If we neglect the volume interaction of the fringe particles, i.e., if we put $\nu(\psi) = \psi$ in the fringe, then we obtain at x > 0

$$\psi(x)=\bar{\psi}\exp\left\{-x\frac{\sqrt{\Lambda-1}}{a}\right\}-[\psi(-0)-\psi(+0)]\alpha_2(x). \tag{4.11}$$

Similarly, if the density n_- is close to the density n_0 (4.2) in the interior of the globule, then at x < 0 (inside the globule) we obtain

$$\psi(x)=\psi_0-(\psi_0-\bar{\psi})\exp\left\{\frac{x}{a}\left[-\frac{2T\varkappa n_0^2+2}{2T\varkappa n_0^2+1}\right]^{1/2}\right\}+[\psi(-0)-\psi(+0)]\alpha_2(-x), \tag{4.12}$$

where $\kappa = -n_0^{-2}(\partial p/\partial n_0)^{-1} < 0$ is the isothermal compressibility of the system of disconnected links with density n_0. In this approximation we have, according to (3.3), for the limiting values of the density

$$\begin{gathered} n_+=\exp\{\tilde{\mu}/T\}, \\ (n_--n_0)/n_0=(\mu(n_0)-\tilde{\mu})\varkappa n_0^2 \end{gathered} \tag{4.13}$$

and, according to (4.5), for the limiting values of the ψ function

$$\psi(\pm 0)=n_\pm\exp\{-\tilde{\mu}/2T\}.$$

This enables us to express the lower limits of integration in (4.9) in terms of $\tilde{\mu}$. For the limiting value of the chemical potential $\tilde{\mu}$ itself we obtain in this approximation the condition[5]:

$$n_+/n_0=-(2T\varkappa n_0^2+1)/(8T\varkappa n_0^2+5). \tag{4.14}$$

We now use the obtained solution to investigate the behavior of the thermodynamic quantities near the coil-globule phase-transition temperature.

For a dense globule to be produced it is necessary that the globule be thermodynamically more favored than the loose coil with F = 0. According to (4.3), this condition reduces to the simple requiremet

$$\mu^*(n_0, T)<0, \tag{4.15}$$

which denotes simply that the self-consistent field should form a potential well. The phase-transition temperature T_0 is thus determined from the conditions

$$\mu^*(n_0, T_0)=0, \quad p^*(n_0, T_0)=0. \tag{4.16}$$

Near the transition point, when $0 < \tau = T_0 - T \ll T_0$, the free energy depends on τ linearly, i.e., the coil-globule phase transition is a first-order transition; it is represented by the line A on the phase diagram (Fig. 3).

To determine the behavior of the surface tension near the transition point it is necessary to take into account the volume interaction of the fringe particles. Since the density profile in the fringe was obtained by us in the zero-order approximation in the interaction, we should, when estimating the interaction itself, first take only pair collisions into account, and, second, assume that the pair-collision probability is not perturbed by the interaction.

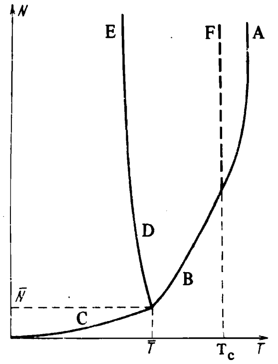

FIG. 3. Phase diagram for a model of "interacting beads on a flexible filament." The lines EDC, EDBA, and CBA delimit the regions of a globule without a fringe, a globule with a fringe, and a coil, respectively.

Since the number of particles in the fringe is vanishingly small in comparison with the total number of the chain particles, it is clear that the fringe consists of loops separated by long sections of the chain lying inside the globular core. Therefore the motions of particles of different loops are practically uncorrelated, and since each particle collides mainly with particles of other loops, the boundary-collision probability is simply equal to $n^2(x)$ (the correction due to collisions with particles of its own loop is of the relative order of magnitude $\sim N^{-1/3}$). Consequently, the energy of the volume interaction of the fringe particles is

$$4\pi R_0^2 TB(T)\int_0^\infty n^2(x)\,dx,$$

where B(T) is the second virial coefficient in the system of disconnected links.[6]

This energy has the same local form as (2.2). Therefore the corresponding refinement of the density profile in the fringe is contained automatically in (4.10), where, however, it is necessary to use in the fringe region the virial expansion for $v(\psi)$ ($v(\psi) \approx \psi(1 + 2B\psi^2)$).

To find the surface-tension coefficient it is obviously necessary to put in (4.4), when integrating over the fringe region x > 0,

$$p^*(n, T)=n^2B(T).$$

When the transition temperature is approached without restriction, we obtain

$$d^2\alpha/dT^2\sim c/\sqrt{T_0-T}+\text{const}. \tag{4.17}$$

Depending on the concrete form of the lateral interaction, the temperature T_0 of the coil-globule phase transition can be either larger or smaller than the critical temperature T_c of the disconnected links. If $T_0 > T_c$, then the surface layer of the globule undergoes a second-order phase transition on going through the temperature T_c; in this transition the distribution of the density becomes discontinuous (dashed line E on the phase diagram of Fig. 3).

5. STRUCTURE OF SMALL GLOBULE WITH COIL

The core of the globule is regarded as small if its volume is $V_0 \ll a^3$. Regardless of its smallness, this core can turn out to be macroscopic if the number of particles in it is large: $n_0V_0 \gg 1$. A globule with a small core can obviously be made up only from a sufficiently short chain, for which the number of particles is $N \ll n_0a^3$.

In view of the smallness of the core (in the absence of an external field), the density everywhere in the interior of the core is approximately constant and equal to its limiting value (in other words, $n_0 \approx n_-$). Therefore, for an extremely small globule we have

$$\hat{g}_{in}(\psi_{in}-\psi_{out})=V_0\{\psi_{in}(0)-\psi_{out}(0)\}g(\mathbf{x})\equiv V_0(\psi_--\psi_+)g(\mathbf{x}).$$

The function in the right-hand side of (2.7), as already mentioned, can vary significantly only over distances ~a, and consequently it is approximately equal to its limiting value in the interior of the core. Since, furthermore, we neglect the volume interaction outside the core

($\mu(n) = T \ln n$ or $v(\psi) = \psi$ at $\psi < \psi_+$), the right-hand side of (2.7) can simply be replaced by $\psi_{out}(\mathbf{x})$:

$$\hat{g}\psi_{out} + V_0(\psi_- - \psi_+)\, g(\mathbf{x}) = \Lambda\psi_{out}(\mathbf{x})$$

(we recall that we are considering in this section the case $\varphi = 0$). The equation can now be solved easily with the aid of a Fourier transformation, and the result is

$$\psi_{out}(\mathbf{x}) = \frac{V_0(\psi_- - \psi_+)}{(2\pi)^3} \int e^{-i\mathbf{k}\mathbf{x}} \frac{g_{\mathbf{k}}\, d^3k}{\Lambda - g_{\mathbf{k}}}, \qquad (5.1)$$

where g_k is the Fourier transform of the kernel $g(\mathbf{x})$.

The constants V_0 and Λ are determined from the boundary condition $\psi_{out}(0) = \psi_+$ of the solution (5.1) and the normalization condition (2.6). Noting that in the assumed approximation we have $n(\mathbf{x}) = \psi^2_{out}(\mathbf{x})$ outside the globular core, and integrating, we obtain

$$NQ(\Lambda) = n_+ a^3, \qquad (5.2)$$
$$V_0(n_- - n_+)/N = P(\Lambda), \qquad (5.3)$$

where

$$J_p(\Lambda) = \frac{a^3}{(2\pi)^3} \int \left(\frac{g_{\mathbf{k}}}{\Lambda - g_{\mathbf{k}}}\right)^p d^3k,$$

$$Q(\Lambda) = J_1^2/(J_1 + J_2), \quad P(\Lambda) = J_1/(J_1 + J_2).$$

To understand the meaning of the solution (5.1) and to use physical intuition, it is expedient to proceed in the following manner. We represent the integrand in (5.1) as a geometric progression and note that the inverse Fourier transform of g_k^m is the convolution of m correlation functions

$$\rho_m(\mathbf{x}) = \int g(\mathbf{x}_1 - 0) \ldots g(\mathbf{x} - \mathbf{x}_{m-1})\, d^3x_1 \ldots d^3x_{m-1} = \frac{1}{(2\pi)^3} \int e^{i\mathbf{k}\mathbf{x}} g_{\mathbf{k}}{}^m\, d^3k.$$

The solution then takes the form

$$\psi_{out}(\mathbf{x}) = V_0(\psi_- - \psi_+) \sum_{m=1}^{\infty} \exp\left\{\frac{m\lambda}{T}\right\} \rho_m(\mathbf{x}). \qquad (5.4)$$

This result could have been guessed directly: the distribution function of the terminal link is $\rho_m(\mathbf{x})$ if a "tail" of m particles is let out into the fringe. On the other hand, the probability of formation of an m-particle "tail" is proportional, according to (2.8), to $\exp\{m\lambda/T\}$. Finally, squaring the series (5.4) (in accordance with the fact that $n_{out} = \psi^2_{out}$), we obtain a clear idea of how the fringe is made up of loops of different lengths. It is useful to bear the expansion (5.4) and the corresponding representations of the integrals J_p in mind in the course of the calculations.

To determine the densities n_+ and n_- we must now minimize additionally the free energy (3.5). With an aim at comparing with a globule without a fringe, we take into account the contribution made to the free energy (2.8) by the volume interaction of the fringe particles. Since the density profile was obtained by us in the zero-order approximation in the interaction, it follows that when the interaction itself is determined only pair collisions need be taken into account. The calculations can be carried through to conclusion in the limiting case $n_+ a^3 \ll NQ_{max}$[7], i.e., $\Lambda \to \infty$. According to (5.4), this corresponds to a fringe consisting only of very short loops. In this case the main contribution to the volume interaction of the fringe particles is made by collision of particles from two different "single-link" loops. The local pressure $p(n_-)$ at the boundary of the globular core should be small in this case, because the number of particles in the fringe is small and the fringe is incapable of experiencing a noticeable mechanical force, although the longitudinal bonds do distribute this force over the entire thickness of the fringe. In other words, the density n_- in the core of the globule is close to the density n_0 determined by the condition $p(n_0, T) = 0$ (cf. (6.2) below). Carrying out additional minimization of the free energy with allowance for all the foregoing, we obtain

$$(n_- - n_0)/n_0 = 2B(T)\, G_4 n_+^2/N, \qquad (5.5)$$

where

$$G_n = \frac{1}{g^n(0)} \int g^n(\mathbf{x})\, d^3x,$$

B(T) is the second virial coefficient in the system of disconnected links. The density n_+ in the fringe is determined by the condition for the continuity of the local chemical potential (3.3) ($n_+ \approx \exp\{\mu(n_0, T)/T\}$). This yields for the equilibrium value of the free energy

$$F = -NT \ln \frac{Ng(0)}{e} + N \frac{f(n_0, T)}{n_0} - n_+ T G_2 - n_+^2 T B G_4 \qquad (5.6)$$

(for a Gaussian kernel $g_G(\mathbf{x})$ we have $G_2 = (2\pi)^{3/2} a^3$, $G_4 = \pi^{3/2} a^3$).

The condition under which this globule is thermodynamically favored over a coil takes the form $F < 0$. Obviously, near the transition point T_0 ($F(T_0) = 0$) the free energy (5.6) goes to zero linearly, i.e., the denaturation of the globule, or conversely, its formation from a coil, is a first-order phase transition, as expected. This transition is represented by line B on the phase diagram (Fig. 3).

6. GLOBULE WITHOUT FRINGE AND PHASE DIAGRAM

We recall that to obtain the entropy of a macroscopic state specified by a density distribution it is necessary to consider a chain without volume interaction in an external field such that the specified distribution is at equilibrium in the field. As shown in Sec. 3, a globule without a fringe is produced when the longitudinal bonds do not make it possible to satisfy the requirement of continuity of the local chemical potential on the boundary of the globule core. Therefore a chain without volume interaction can form a globule without a fringe only if the external field outside the globule becomes infinite. This means that a globule without a fringe is described by Eqs. (2.2)–(2.8), in which the operator $\hat{g}$ is replaced by the operator $\hat{g}_{in}$. In particular, the equilibrium density distribution is determined from the equations

$$\hat{g}_{in}\psi = \psi \exp\left\{\frac{\mu^*(n) - \lambda}{T}\right\}, \qquad (6.1)$$
$$\Lambda n = \psi \hat{g}_{in} \psi.$$

The limiting value of the density is determined by additional minimization of the free energy (2.8).

The description of the globule without a fringe is particularly simple if the fringe has extremely small dimensions. The density in a small globule is practically constant, so that Eq. (6.1) can easily be solved:

$$g(0) V_0 \psi(0) = \psi(0) \exp\left\{\frac{1}{T}[\mu^*(n_0, T) - \lambda]\right\},$$

after which the additional minimization of the free energy yields the following condition: for the density n_0 in the globule:

$$p(n_0, T)=0. \tag{6.2}$$

The meaning of this result is explained by the simple consideration that in a small globule without a fringe all the particles are contained in a small volume (much smaller than a^3), and therefore the mechanical equilibrium of the boundary cannot be due to the tension of the longitudinal bonds (cf. (5.5)).

For the equilibrium value of the free energy of a small globule without a fringe we obtain

$$F=-NT\ln\frac{Ng(0)}{e}+N\frac{f(n_0,T)}{n_0}, \tag{6.3}$$

so that the condition under which the globule without a fringe is favored thermodynamically over the coil ($F < 0$) takes the form

$$N>\frac{1}{g(0)}\exp\left\{\frac{\mu(n_0,T)}{T}+1\right\}. \tag{6.4}$$

For a Gaussian kernel we have $1/g_G(0) = (4\pi)^{3/2}a^3$. The corresponding first-order phase transition is represented by curve C on the phase diagram (Fig. 3).

Comparison of the free energies of small globules without fringes (6.3) and with fringes (for which purpose it is precisely the expansion (5.4) and its corollary (5.6) that are well suited) leads to the following condition under which the fringe is not favored:

$$-B(T)n_+>G_2/G_4 \tag{6.5}$$

($n_+ \approx \exp\{u(n_0, T)/T\}$; for a Gaussian kernel we have $G_2/G_4 = 2\sqrt{2}$). We recall that at low temperatures the second virial coefficient behaves like $B(T) \sim -\exp\{-U_0/T\}$, where $U_0 < 0$ is the characteristic energy of the lateral attraction between links.

A phase transition in which the fringe is "hidden" in the globule is obviously a first-order transition. For small globules, it occurs when the equality is reached in (6.5). Refinement of (6.5) by taking into account the next higher terms in the expansion (5.4) reveals a weak dependence of the transition temperature on the number of links. This phase transition is represented by curve D on the phase diagram.

It is likewise thermodynamically more convenient for an extremely long chain to form, at sufficiently low temperature, a globule without a fringe. The distribution of the density in the globule can be obtained by approximating the integral operator $\hat{g}_{in}$ by a differential operator, in analogy with the procedure used for the operator $\hat{g}$ in Sec. 4. The expressions for the resultant unfavorable conditions for fringe formation are quite complicated, because the difference between the free energies of large globules with and without fringes is of the order of the surface energy. For an extremely long chain, the temperature at which the fringe vanishes does not depend on the length of the chain, and this is represented on the phase diagram by the straight line E.

The position of the triple point on the phase diagram (we designate its coordinates by $\overline{N}$ and $\overline{T}$) are determined by the fact that the free energies of the globules, with and without the fringe, should vanish at this point. We see immediately, however, that the triple point lies in the region where the small-globule approximation does not hold ($\overline{N} \ll n_0a^3$). Its coordinates are therefore determined by the equations

$$eG_2/G_4g(0)=-B(\overline{T})\overline{N},\quad \overline{N}g(0)=\exp\{\mu(n_0,\overline{T})/\overline{T}+1\}. \tag{6.6}$$

Since Eq. (6.6) has a solution only at a sufficiently low temperature, and in particular at a temperature lower than the critical temperature T_c in the system of disconnected links, the triple-point temperature is certainly lower than T_c: $\overline{T} < T_c$.

Putting

$$N=\overline{N}+\delta N\quad(\delta N\ll\overline{N}),$$
$$T=\overline{T}+\delta T\quad(\delta T\ll\overline{T}),$$

we obtain equations for all three equilibrium curves near the triple point:

1. The equilibrium curve B (coil-globule with fringe:

$$\left\{\frac{1}{n_0}\left[\frac{\partial f(n_0,T)}{\partial T}-\frac{f(n_0,T)}{T}\right]\left[1+\frac{n_+}{\overline{N}}G_2\right]+G_2\frac{Tn_+}{B(T)}\frac{\partial B}{\partial T}\right\}_{T=\overline{T}}\frac{\delta T}{\overline{T}}=\frac{\delta N}{\overline{N}}.$$

2. Equilibrium condition C (coil-globule without fringe):

$$\left\{\frac{1}{n_0}\left[\frac{\partial f(n_0,T)}{\partial T}-\frac{f(n_0,T)}{T}\right]\right\}_{T=\overline{T}}\frac{\delta T}{\overline{T}}=\frac{\delta N}{\overline{N}}.$$

3. Equilibrium condition D (globule with fringe-globule without fringe):

$$\delta T=0.$$

Finally, to illustrate all the foregoing, we can plot the fraction of the particles in the globular core against the temperature. This plot is shown in Fig. 4 for a chain with more than N links.

7. FORMULATION OF SMALL GLOBULE IN THE PRESENCE OF AN EXTERNAL FIELD

In the case of a small globule, it is preferable to obtain the solution by assuming that both an external field $\varphi(\mathbf{x})$ and volume interaction $\mu^*(n)$ are simultaneously present, since a potential field localized in a small volume can, for example, correspond to interaction of the polymer links with any particular particle. If the globule is formed by the "joint efforts" of the external field and the self-consistent field, then the regions in which each of them differs from zero do not in general coincide at all. Let us assume, however, that both fields are equal to zero outside the volume V, and assume furthermore that this volume is small: $V \ll a^3$. Then the longitudinal bonds inside this line have no effect whatever on the density distribution: according to (3.2), the total local chemical potential $\mu(n) + \varphi(\mathbf{x})$ inside the volume is practically constant and equal to its external limiting value:

$$\mu(n(\mathbf{x}))+\varphi(\mathbf{x})=T\ln n_+\ \text{(inside V)}. \tag{7.1}$$

this yields, in accord with (3.1),

$$n(\mathbf{x})=\psi_+\psi(\mathbf{x})\quad\text{(inside V)} \tag{7.2}$$

(n_+ and ψ_+ are the values of $n(\mathbf{x})$ and $\psi(\mathbf{x})$ on the outer boundary of the volume V; $n_+ = \psi_+^2$).

Unlike the case analyzed in Sec. 5, when there was no

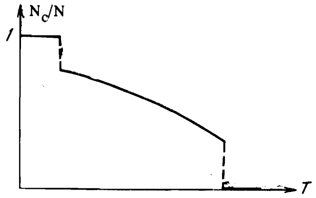

FIG. 4. N_c is the number of particles in the core of the globule; N is the total number of particles in the chain.

external field, the density and the ψ function inside the volume V are not constant. Nonetheless, it can be verified that the ψ function $\psi_{\text{out}}(\mathbf{x})$ outside V can be obtained in essence in the same manner as in Sec. 5. The resultant formulas differ from (5.1)–(5.3) only in that ψ_- and n_- are replaced throughout by the corresponding mean values

$$\langle\psi\rangle_-=\frac{1}{V}\int\limits_V \psi_{in}(\mathbf{x})\,d^3x,$$

$$\langle n\rangle_-=\frac{1}{V}\int\limits_V n(\mathbf{x})\,d^3x=\psi_+\langle\psi\rangle_-.$$

If the self-consistent field of the volume interaction does not form an additional discontinuity of the density distribution, then these equations account for the problem completely. On the other hand, if such a discontinuity does occur on some surface (the temperature should in this case be lower than T_c and there should be no excessive compression on the part of the external field), then the value of the chemical potential on this surface should be determined by additional minimization of the free energy.

For example, if the globule is formed around a particle and the interaction of the monomers with this particle consists only in that they cannot penetrate into the particle (φ is equal to $+\infty$ inside the particle and to zero outside the particle) then, as expected, the density of the globular core around this particle is the same as the density of the free globular core, i.e., it is determined by (5.5) (of course, it is thermodynamically not profitable for the globule to absorb this particle, but this is already a purely surface effect).

We now proceed to analyze the phase transition. The condition for the thermodynamic profitability of the globular state takes the form $F < 0$, and we should investigate the dependence of the free energy on the temperature near the transition temperature, i.e., at $0 < \tau = T_0 - T \ll T_0$.

For a chain without volume interaction we have

$$F=N\lambda=-NT\ln\Lambda,$$

so that the phase transition occurs as $\Lambda \rightarrow 1$. Since $P(\Lambda) \sim \sqrt{\Lambda-1}$ as $\Lambda \rightarrow 1$, we have $F = c\tau^2$ at small τ, i.e., the coil-globule transition is of second order. Mathematically, this transition is due to the appearance of a discrete eigenvalue in (2.7). By the same token, the transition point is singular for each of the phases. If the volume interaction is present but is small, so that the transition is due mainly to the action of the external field, then, owing to the second term of (2.8), the transition occurs at a value of Λ somewhat larger than unity, and becomes, strictly speaking, a first-order transition $F = b\tau + c\tau^2$, but the coefficient b is small to the extent that the volume interaction is small. If there is no external field at all, then we obtain a genuine first-order phase transition, and the transition point, as for all first-order transitions, is by itself not singular in any way. Finally, if both the external field and the volume interaction are not small, then two different first-order phase transitions will occur at different temperatures.

The density distribution in a small globule without a fringe is determined in the presence of an external field by the condition that the chemical potential be constant. It must be remembered, however, that the continuity of the chemical potential is violated on the very boundary of a globule without a fringe.

8. PHASE DIAGRAM OF A REAL POLYMER GLOBULE

Figure 3 shows the phase diagram for the "beads-on-a string" model chain. Analytic expressions for all the equilibrium curves of this diagram were obtained in the preceding sections. Of course, in a real chain there is interaction not only between isolated beads, but also to a greater or lesser degree between all parts of the chain. Therefore the actual applicability of the derived formulas is limited to the case of a heteropolymer chain with sufficiently flexible primary core and to loose monomers, which interact with one another much more strongly than the remaining sections of the chain. Nonetheless, the qualitative conclusion that three different macroscopic phases exist (coil, globule with fringe, globule without fringe) is itself valid for a real chain. It is also obvious that the transitions between these phases are of first order. A rough approximation of the corresponding phase diagram is shown in Fig. 5.

The shape of the resultant globule is determined by the character of the primary structure of the chain. If the chain is very long and its primary structure is statistically uniform to a sufficient degree, then the globule is close to spherical. For a shorter chain, the inhomogeneity of the primary structure becomes important and the chain can form a globule of complicated shape. In particular, if the chain consists of several pieces, each capable of forming a globule, then a polyglobule consisting of several subunits is produced. The conclusion that the fringe vanishes remains in force in this case, too.

Obviously, our entire analysis is valid only at temperatures at which the solvent (water) is in the liquid state. The relation between the melting temperature of ice and the boiling temperature of water on the one hand and the characteristic temperature points of the phase diagram on the other depends on the concrete properties of the chain and can, generally speaking, be arbitrary.

9. MICROSTRUCTURE OF THE GLOBULE CORE AND POSSIBILITY OF ITS SELF-ORGANIZATION

For sufficiently large homopolymer globules, the microscopic structure of the globular core should differ little from the equilibrium structure of the equivalent homogeneous phase without linear memory but with the same density n_0. This qualitatively obvious circumstance is reflected in the results of Sec. 4, namely, by virtue of (4.2) the condition (4.15) under which the globular state is thermodynamically favored, takes the form $f(n_0, T) < f_{id}(n_0, T)$, i.e., it coincides with the condition under which a homogeneous condensed phase is thermodynamically favored in a system of "beads" without a linear memory.[8] Consequently, the core of such a globule can

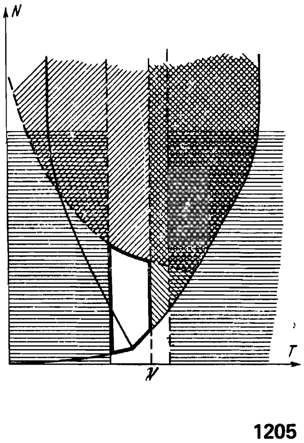

FIG. 5. Approximate form of the phase diagram of a real heteropolymer globule. The shaded regions are those where self-organization of the tertiary structure is impossible for one reason or another: horizontal shading–the solvent is in the solid or gaseous state; shading with lines inclined to the right–large probability of local topological irregularities on the surface of the globule or in its interior); shading inclined to the left-thermodynamically favored restructuring of the tertiary structure as a whole.

be similar to a "liquid" or to a "crystal." When speaking of a liquid-like structure, we have in mind not the mobility of its elements, but the disorder or amorphous character of this structure, by virtue of which it can be realized with the aid of a tremendous number of essentially different conformations. To the contrary, when speaking of a crystal-like structure, we should pay principal attention not to the fact of rigorous ordering, but the fact that this ordering corresponds to spatial periodicity.

If we mean by crystal-like structure an aperiodic rigorously ordered structure, which is realized, apart from thermal vibrations or deformations, with the aid of a single "optimal" conformation, then a similar subdivision of equilibrium structures can also be used for heteropolymers with arbitrary sequence of the chain links. From this point of view, the possibility of self-organization of the tertiary structure means that the equilibrium globule should have a crystal-like structure. Let us ascertain when this is possible.

We note first that the chain can have an entire set of utterly different spatial arrangements (conformations) corresponding to one and the same value of the "energy" E. This "energy" includes both the chain energy and the entropy relative to thermal vibrations or other deformations that do not change the general character of the tertiary structure. We denote by E_0 the minimal value of the "energy" realized by the "optimal" conformation of the chain. We emphasize that we are dealing with extensive restructuring of the chain, for which the "excitation" $E - E_0$ is of the order of the volume energy, $E - E_0 = N\epsilon$. The logarithm of the number of macroscopic conformations realizing the "energy" E is the conformation entropy S, which is a function of E: $S = S(E)$. For a heteropolymer we need not expect the degeneracy at which the derivative $\partial S/\partial E$ would vanish at the "energy" E_0. Therefore at small $E - E_0$ we have $S(E) = (E - E_0)/\gamma = N\epsilon/\gamma$, where the coefficient γ is determined, in order of magnitude, by the energy of the lateral interactions per particle. The total free energy is consequently $\mathscr{F} = N\epsilon(1 - T/\gamma)$. We see that at a relatively high temperature ($T > \gamma$) the thermodynamically favored state is a mixture of a set of essentially different conformation states, i.e., the resultant tertiary structure is liquid-like.

Thus, in order for a crystal-like spatial structure to be produced it is necessary above all to satisfy the condition $T < \gamma$. However, even when this condition is satisfied, local violations of order can also occur in the system and change the entropy (per defect) by an amount $\sim \ln c$, where c is the concentration of the local defects. In an ordinary crystalline system, the local violation of the ordering (say an interchange of a pair of neighboring particles), which is always present as a result of the thermal motion and is proportional to its activation energy, still does not disrupt the crystal structure. In a polymer globule, such a local violation is most frequently connected with topological changes in the distribution of the segments of the chain in the vicinity of the local disruption; it therefore calls for an essential change in the conformation as a whole. Naturally, a rewinding of the chain cannot occur in the already produced globule, but a large probability of such an event would mean that the equilibrium ensemble of globules constitutes a mixture of systems with different spatial organization, i.e., there is no steady-state tertiary structure. The characteristic energy U_D of such a defect can be essentially different, depending on whether the defect is in the interior of the globule ($U_D^{(1)}$) of on its surface ($U_D^{(2)}$). Inside the globule, the minimal defect concentration is $\sim N^{-1}$, and on the surface this minimal concentration is $\sim N^{-2/3}$. Therefore the "volume" and "surface" topological disturbances change the free energy by amounts $\sim(U_D^{(1)} - T \ln N)$ and $\sim(U_D^{(2)} - 2/3\ T \ln N)$, respectively. It follows, therefore, that the requirement that the system be stable against local disturbances imposes a stringent requirement on the chain length, $N \ll N_c$, where the critical chain length is $N_c \sim \exp\{U_D/T\}$.

The energy of a topological defect can be very small (for example, much smaller than the energy of one hydrogen bond) if the globule is not compact enough. Particularly dangerous in this sense are surface disturbances. Therefore compactness of the globule and its possession of a sufficiently large surface tension are essential if self-organization of its crystal-like structure is to be possible. This imposes very stringent limitations on the primary structure.

In the case of proteins, a limitation of this type, in particular, is the relation noted by Fisher[2] between the concentration of the hydrophilic links and the dimensions and shape of the globule, and also additional correlations in the arrangements of these links.

In a more concrete analysis of the tertiary structure of protein globules, an important role is assumed by the presence of the secondary structure, which we disregarded completely in the analysis of an arbitrary polymer chain (see[3]). We note in this connection that sections of kinks, where the helical secondary structure is violated, are most dangerous in the sense of the probability of local disturbances.

An alternative to the discussed "equilibrium" character of the tertiary structure of the globules is the kinetic mechanism of its formation. This is equivalent to the assumption that the "optimal" conformation is not determined by the absolute minimum of the free energy, but corresponds kinetically to the most rapidly attainable relative minimum. For a long chain, the fastest organization takes place if it can occur independently in different sections of the chain. This is possible only in the case when the resultant spatial structure has the simplest topology, i.e., it has a minimal (zero) number of nodes (or quasinodes[9]).

It appears, however, that for reliable self-organization of a tertiary structure in a long chain it is necessary that the kinetic and thermodynamic requirements coincide, i.e., that the energywise optimal conformation have the simplest topology. (In accordance with the statements made in Sec. 8, a polymer chain with a corresponding primary sequence can form a polyglobular structure that is compatible with both requirements.)

Indeed, it is obvious that the self-organization of a conformation that is kinetically difficult to achieve and is thermodynamically optimal is not reliable if this conformation is not kinetically optimal. On the other hand, however, if the kinetically-optimal configuration does not coincide with the thermodynamic minimum, then it cannot occur at all in the vicinity of the true transition temperature, and the more difficultly attainable thermodynamic minimum must be realized. This contradiction makes the foregoing statement convincing (we emphasize that we are dealing with organization of the spatial structure of a ready-made chain, for example, after denaturation of the globule, and not with organization during the course of synthesis on the ribosome).

Inasmuch as the overwhelming fraction of the conformations corresponds to strongly knotted structures with increasing chain length, this requirement imposes an extremely stringent limitation on the character of the admissible sequences. In other words, only a negligibly small fraction of all the possible primary structures is capable of biochemical functioning. It is precisely these structures which were selected during the process of prebiological and biological evolution.

APPENDIX

THE CONCEPTS OF "PARTIAL EQUILIBRIUM" AND "SMOOTHED DENSITY"

Partial equilibrium of a macroscopic system sets in in those cases when the probability distribution of some functions of the microscopic state of the system is specified by factors that are external with respect to the system. If the relaxation time of a statistical system with respect to a certain parameter is much larger than with respect to all other parameters, then one can speak of a state of partial equilibrium (or incomplete relaxation), in which the value of the considered parameters is set from the outside (by the initial conditions), and the system is in equilibrium with respect to all the remaining variables. Naturally, this approximation is valid only at times that are short in comparison with the total system relaxation time. The method of describing a state of partial equilibrium consists in choosing the regions of integration in phase space and ascribing suitable weights to them.

The relaxation time of a polymer chain relative to breaking (i.e., relative to polymerization–depolymerization processes) is much longer than the relaxation time relative to kinks. Therefore, in the study of conformations of a polymer chain we should regard the thermodynamic equilibrium of the chain in a solution as a state of partial equilibrium with a fixed sequence of the chain links. This state of partial equilibrium is precisely the one described by the distribution (2.1).

The macroscopic state of a chain is defined as a state of partial equilibrium with a fixed distribution of the smoothed density. The smoothed density can be defined, for example, in the following manner. We break up all of space into small cubes of volume v, and assume the "smoothed" density inside each cube to be constant and equal to the average value of the density in this cube. To be sure, the application of the word "smoothed" to the distribution obtained in this manner is quite arbitrary, since such a "smoothed" density takes the form of a piecewise-constant function with steps. To avoid this, we define the smoothed density $n_G(\mathbf{x})$ at the point x as the mean value of the exact density function in a sphere of volume v around the point x:

$$n_\Gamma(\mathbf{x})=\frac{1}{v}\int_v\sum_j\delta(\mathbf{x}+\mathbf{r}-\mathbf{x}_j)\,d^3r.$$

Any of these or similar methods of defining the smoothed density leads, of course, to perfectly equivalent results. The convenience of the last definition lies in the fact that the smoothing volume v need not necessarily be the same for all the points x. It is only necessary that the number of particles in each smoothing element be sufficiently large ($nv \gg 1$), and then the specification of a smoothed density denotes the fixing of only a relatively small number of variables $N/nv \ll N$ (it is precisely in this case that a state with a fixed smoothed density can be called a macroscopic state). Such a smoothed density corresponds to a large volume in configuration space. The logarithm of this volume is the configuration entropy of the macroscopic state:

$$S\{n(\mathbf{x})\}=\ln\int\delta(n_\Gamma(\mathbf{x})-n(\mathbf{x}))\prod_j g_j\,d\Gamma.$$

The free energy of the described macroscopic state is

$$\mathscr{F}\{n(\mathbf{x})\}=-T\ln\int\exp\left\{-\frac{\varepsilon(\Gamma)}{T}\right\}\delta(n_\Gamma(\mathbf{x})-n(\mathbf{x}))\prod_j g_j\,d\Gamma.$$

In a thermodynamic description of the partial-equilibrium state, the quantity

$$E\{n\}=\mathscr{F}\{n\}+TS\{n\}$$

plays the role of the energy, although actually, of course, it does not coincide at all with the true energy of the system.

The free total-equilibrium energy[10)] is obviously equal to

$$F=-T\ln\int\exp\{-\mathscr{F}\{n\}/T\}Dn,$$

where the integration is carried out over all the possible distributions of the smoothed density. Actually, however, the integration is only with respect to N/nv variables, which determine the smoothed density (this is particularly clear for the first of the indicated methods of defining the smoothed density). Since the number of particles in the smoothing elemet is large, the integral can be determined by the saddle-point method, and we obtain as a result

$$F=\min\{\mathscr{F}\{n\}\}=\mathscr{F}\{\bar{n}\},$$

where $\bar{n}$ is the equilibrium distribution of the smoothed density. The saddle-point method introduces an error $(\ln nv)/nv$. Since most particles in the element v pertain to remote sections of the chain, natural fluctuations $\sim(nv)^{-1/2}$ take place in this element and are much larger than the error of the saddle-point method (in exactly the same manner as in an ordinary system without linear memory under the analogous situation).

We recall that the Boltzmann entropy

$$S=-\int n(\mathbf{x})\ln\frac{n(\mathbf{x})}{e}\,d^3x$$

depends on a density that is smoothed out in the same sense.

[1)] We normalize the correlation function by the condition $\int g(y)d^3y = 1$.

[2)] Since the correlation function g between the positions of the ends of such a composite "link" is made up by convolution of several initial correlation functions, it begins to exhibit the influence of the central limit theorem of probability theory and, consequently, should differ little from the Gaussian function

$$g_\Gamma(\mathbf{x})=(4\pi a^2)^{-3/2}\exp\{-x^2/4a^2\},$$

which can therefore be usefully regarded as the model of the correlation function.

[3)] Here and throughout, when speaking of the free energy, the chemical potential, etc., we have in mind the configurational parts of these quantities.

[4)] Expressions (2.2) and (2.3) were written down by us in such a way that the free energy is reckoned from the level of the free energy of the free coil.

[5)] The conditions under which the radicand in (4.12) and in the right-hand side of (4.14) is positive are connected with the conditions under which the integral operator $\hat{g}$ can be approximated by a differential operator. If these conditions are not satisfied, then the surface layer of the globule can have a complicated superstructure connected with the anisotropy of the density distribution. The characteristic dimension of this superstructure should be ~a. This case is not considered in the present paper.

(6) It is given by the expression

$$B(T)=\frac{1}{2}\int\left\{1-\exp\left\{-\frac{U(\xi)}{T}\right\}\right\}d^3\xi$$

where U(ξ) is the potential of the "bead–bead" pair interaction.

(7) For purely arithmetic reasons, the maximum of Q can differ from unity in order of magnitude. For example, for a Gaussian kernel, numerical integration yields $Q_{\max} \approx 3 \times 10^{-3}$.

(8) We recall that the free energies of large globules with and without fringes differ only in the value of the surface energy, so that the phase transition connected with the vanishing of the fringe produces practically no change in the microstructure of the internal part of the globular core.

(9) Strictly speaking, the concept of a node exists only for a closed chain. However, for very long chains, with length L much larger than the globule dimension R ($L \gg R$), it is always possible to "close" the chain by joining its ends with the aid of a suitably placed segment of small length ~R and by the same token introduce the concept of quasinodes.

(10) We have in mind total equilibrium relative to all possible distributions of the smoothed density, i.e., relative to kinks in the chain, and not to breaks.

---

[1] I. M. Lifshitz, Zh. Eksp. Teor. Fiz. 55, 2408 (1968). [Sov. Phys.-JETP 28, 1280 (1969)].
[2] H. Fisher, Proc. Nat. Acad. Sci. USA 51, 1285 (1964).
[3] O. B. Ptitsyn, Uspekhi Sovrem. Biol. 69, 26 (1970).

Translated by J. G. Adashko
250