SOME PROBLEMS OF THE STATISTICAL THEORY OF BIOPOLYMERS

I. M. LIFSHITZ

Physicotechnical Institute, Ukrainian Academy of Sciences

Submitted July 9, 1968

The basis of any biological structure is the extremely long molecular chains of biopolymers, whose links are joined together with strong valence bonds. However, in addition to these bonds, there are also other forces of "lateral" or "volume" interactions between the links of the chain that have approached one another because of its bending. They determine the phase state of the system that arises. In the absence of a volume interaction, a free chain winds itself into a loose random coil, owing to fluctuations in the bending. At certain critical values of the temperature and other parameters of the medium, additional attractive forces lead to collapse of the loose random coil into a compact globule. Such globular states are essentially states of partial equilibrium in which the primary structure of the chain and the bonds between its links are considered to be fixed, and to form a "linear memory" in the system. Equilibrium is attained with respect to the additional interaction forces, with restrictions imposed by the linear memory. A systematic statistical-thermodynamic theory of such systems is constructed in this study, and the random coil-globule transition is analyzed in these systems.

INTRODUCTION

THE basis of any biological structure is the extremely long molecular chains of biopolymers. These chains are suspended in the solvent medium, and occur in various conformational states, depending on the strength and nature of the volume interaction between the elements of the chain, the temperature, and the parameters of the medium. Since each chain contains an enormous number of links, it can be considered as a statistical system, and its different conformational states as the macroscopic phases of this system. In order to construct the statistical thermodynamics and kinetics of biopolymers, we must formulate the characteristic distinguishing features of these systems, which might be introduced into their statistical models to construct a general and systematic theory.

The primary structure of a chain fixes the sequence A_j of its links (identical or differing). The probabilities of the relative positions and orientations of adjacent links are intercorrelated. This primary structure is considered invariant, short of rupture of the polymer chain, and forms a fixed linear memory of the system. Since the characteristic time of conformational relaxation of a long and flexible chain $\tau \sim N^2$, the energy barriers E_0 preventing rupture of the longitudinal valence bonds in the primary chain must satisfy the condition of stability $\omega\tau \ll 1$ (where $\omega \sim e^{-E_0/T}$ is the frequency of chain rupture). Evidently, we thereby have generally a case only of states of partial equilibrium, and the fixed memory is an expression of this fact.

The simplest type of memory involves only fixing the site of each link in the linear sequence along the chain, i.e., the fact that the links seem to be connected together by a continuous flexible cord. This memory exists even in a chain of identical elements (homopolymer), and we shall simply call it the linear memory. It can be described by the correlations g_j between adjacent links. However, in addition to the interactions along the chain,

which are already taken into account by assigning the correlations g_j, there is also a "lateral" or "volume" interaction of its elements. Generally speaking, this is an interaction between distant links of the chain that have come together because of its bending. This is what determines the phase states that arise in the system. If $\epsilon(\Gamma)$ is the energy of this interaction, the linear memory is manifested in the fact that the equilibrium distribution function $e^{-\epsilon(\Gamma)/T}$ is replaced by $e^{-\epsilon(\Gamma)/T}\prod_{j=1}^{N} g_j$, while the statistical integral Z takes on the form

$$Z = \int e^{-\epsilon(\Gamma)/T} \prod_j g_j \, d\Gamma.$$

In complex heteropolymers having differing links, the sequence is fixed by the additional informational memory. In principle, we could take into account the concrete and complex sequence of links by detailed tracing of the volume energy interactions arising in such a system. However, this course is practically unrealizable, and most often the informational memory can be taken into account at least partially by introducing "informational interactions," along with the energy interactions. The former amount to extra prohibitions and limitations on the allowable region of integration in phase space (in other words, it amounts to the systematics and selection of allowable microstates).[1] These prohibitions also correspond to the expeditious introduction of states of partial equilibrium.

The considerable variety in the phases that arise involves the nature of the interaction forces between the regions of the chain that have approached one another. Thus, an interaction involving saturation of the free hydrogen bonds of each of the links leads to the combination of two or three chains into a helical cable, such as forms the secondary structure of a biopolymer (DNA,

---

[1] For example, one can conveniently introduce various models of informational interaction in analyzing helix-coil transitions in biopolymers having a helical secondary structure.

fibrous proteins). The volume interaction between the links of a flexible chain entangled into a loose random coil involves weaker forces not generally having the feature of saturability (e.g., van der Waals forces), and leads to the appearance of relatively compact globules of characteristic structure. The nature of the problems arising with regard to helix-coil and coil-globule transitions differs considerably.

We shall discuss in this paper some general problems of statistical thermodynamics of long chains, and problems involving the random coil-globule transition in systems having a simple linear memory (i.e., chains consisting of identical elements and systems reducible to them). While the actual structure of globular proteins with their complex informational memory proves to be incomparably more loose-structured and distinctive, however, the sensitivity of the system to this information resides in the presence of a linear memory, and it becomes evident even from the simplest theory.

We shall begin with establishing some statistical-thermodynamic properties of chains having no volume interaction, but having an external field constraining the system. The concepts introduced here and the obtained results will permit us to formulate and solve the problem of the volume interaction in such systems.

1. DENSITY AND DENSITY CORRELATIONS IN A FREE RANDOM COIL

Let x_j ($j = 1, ..., N$) be the coordinates of the links of the chain (these are generally the coordinates of the links that can interact with one another or with an external field (see Fig. 1)). The distribution function in configuration space for a free chain is given by the correlations between adjacent links $g_j = g(x_j, x_{j+1})$:

$$\rho(x_1, ..., x_N) = \prod_j g_j, \quad g_j = g(y_j), \quad y_j = x_{j+1} - x_j, \quad \int g d^3y = 1. \quad (1.1)$$

(We shall assume for simplicity that g_j depends only on the coordinates of adjacent links, but not on their orientations.) For large k, the probability $\rho_k(x)$ that the k-th link should be located at the point x, when the zeroth link is fixed at the origin, will be

$$\rho_k(x) \sim k^{-3/2} e^{-x^2/4ka^2}; \quad a^2 = \frac{1}{6}\overline{y^2} = \frac{1}{6}\int g(y)y^2 d^3y. \quad (1.2)$$

As we know, when $N \gg 1$, owing to fluctuations in curvature, a free chain will be wound into a random coil of linear dimensions $\sim R_N = a\sqrt{N}$ and of volume
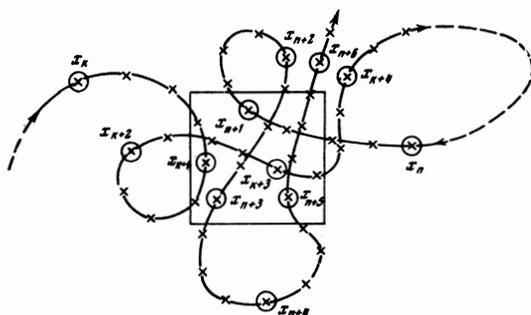

FIG. 1. Schematic drawing of a region in space of a polymer chain. The circles mark those links that can interact substantially with one another or with an external field.

$V_N \sim a^3 N^{3/2}$. Its mean density is: $N/V_N \sim a^{-3}N^{-1/2}$. Let us formulate some properties of such random coils in a form that we shall require hereinafter.

Let us introduce the density of links at the point x:

$$n(x) = \sum_{j=1}^{N} \delta(x - x_j). \quad (1.3)$$

Its mean value

$$\bar{n}(x) = \sum_{j=1}^{N} p_j(x) = N \int_0^1 p_{sN}(x) ds$$

will be, according to (1.2):

$$\bar{n}(x) d^3x = N\rho(\xi) d^3\xi, \quad \xi = x/R_N, \quad R_N = a\sqrt{N},$$

$$\rho(\xi) = \frac{1}{8\pi^{3/2}} \int_0^1 e^{-\xi^2/4s} \frac{ds}{s^{3/2}}. \quad (1.4)$$

On the new scale $\xi = x/R_N$, the density $\rho(\xi)$ differs considerably from zero at distances $\xi \sim 1$. That is, the radius of the system $\xi_N \sim 1$. However, the density correlation

$$\frac{\overline{n_1 n_2} - \bar{n}_1 \bar{n}_2}{\bar{n}_1 \bar{n}_2} = \frac{\rho_{12}(\xi_1, \xi_2)}{\rho(\xi_1)\rho(\xi_2)} - 1,$$

$$\rho_{12}(\xi_1, \xi_2) = (4\pi)^{-3/2} \int_0^1 \int_0^1 \exp\left\{ -\frac{\xi_1^2}{4s_1} - \frac{|\xi_1 - \xi_2|^2}{4|s_1 - s_2|} \right\} \frac{ds_1 ds_2}{s_1^{3/2}|s_1 - s_2|^{3/2}}$$

(1.5)

also declines at distances $\xi \sim 1$. Thus, the correlation radius turns out to be of the order of the radius of the system, and the density ceases to be a thermodynamic parameter (i.e., an exact quantity as $N \to \infty$). This means that the random coil pulsates macroscopically, and the density $\rho(\xi)$ given by Eq. (1.4) is the time-average value. The characteristic time of these macroscopic pulsations (the relaxation time) is $\tau \sim N^2$. However, such a situation arises because the random coil is free, i.e., the external pressure is zero. In other words, there are no external fields constraining the system, nor any volume attraction between the links. The situation changes radically if we include interactions (whereby the density fluctuations can be made arbitrarily small). A free random coil is a degenerate case, and in certain conditions does not give a good zero-order approximation for describing such a system even in the presence of weak interactions (as $N \to \infty$).

Hence, the first problem to solve is to derive a statistical thermodynamics for random coils in an external field constraining the system. Such a derivation will permit us to introduce and define important statistical-thermodynamic concepts for chains, including the entropy of a random coil as a functional of its smoothed density $n(x)$ (when the latter is made a thermodynamic quantity).

2. STATISTICAL THERMODYNAMICS OF A RANDOM COIL IN AN EXTERNAL FIELD

Let us express the energy characteristics in temperature units, and denote the energy of an individual link at the point x by $\varphi(x)$; $\varphi(x) = u(x)/T$. The energy of the system at the point Γ of configuration space is

$$\varepsilon = \sum_{j=1}^{N} \varphi(x_j) = \int n(x)\varphi(x) d^3x. \quad (2.1)$$

The free energy $F\{\varphi\}$ of the random coil in the external field with respect to the free energy level of a free coil will be

$$\frac{\mathcal{F}_T}{T} \equiv F \equiv F\{\varphi\} - F\{0\} = -\ln Z_N, \qquad (2.2)$$

$$Z_N = \int \exp\left\{-\sum \varphi(\mathbf{x}_j)\right\} \prod_j g(\mathbf{y}_j) d\Gamma, \quad d\Gamma = \prod_j d^3x_j = \prod_j d^3y_j. \quad (2.3)$$

The equilibrium density is

$$\bar{n}(\mathbf{x}_j) = \delta F / \delta\varphi. \qquad (2.4)$$

The energy is given by the expression

$$\frac{\varepsilon_T}{T} \equiv E \equiv E\{\varphi\} - E\{0\} = (\varphi, \bar{n}) = \left(\varphi, \frac{\delta F}{\delta\varphi}\right), \qquad (2.5)$$

and the entropy by

$$S = S\{\varphi\} - S\{0\} = E - F = (\varphi, \bar{n}) - F. \qquad (2.6)$$

Here, as usual, (f, g) denotes $\int fg d^3x$.

In order to calculate Z_N, let us introduce the quantity

$$Z\left(\begin{matrix} \mathbf{x}', \mathbf{x} \\ 1, N \end{matrix}\right) = \int \exp\left\{-\sum \varphi(\mathbf{x}_j)\right\} \delta(\mathbf{x}_1 - \mathbf{x}')\delta(\mathbf{x}_N - \mathbf{x}) \prod_j g_j d\Gamma, \quad (2.7)$$

This is the statistical integral for the fixed values $\mathbf{x}_1 = \mathbf{x}'$, $\mathbf{x}_N = \mathbf{x}$. That is, apart from a coefficient, it is the probability of finding the beginning and the end of the chain at the points $\mathbf{x}'$ and $\mathbf{x}$, respectively.

When considered as a function of $\mathbf{x}$ and N, the quantity $Z_N(\mathbf{x}) = Z\left(\begin{matrix} \mathbf{x}', \mathbf{x} \\ 1, N \end{matrix}\right)$ satisfies the obvious recurrence relation

$$Z_{N+1}(\mathbf{x}) = e^{-\varphi(\mathbf{x})}\hat{g}Z_N,$$
$$\hat{g}f = \int g(\mathbf{x} - \mathbf{x}')f(\mathbf{x}')d^3x' = \int g(\mathbf{y})f(\mathbf{x} - \mathbf{y})d^3y \qquad (2.8)$$

with the initial condition $Z_1(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}')e^{-\varphi(\mathbf{x})}$. Its solution is:

$$Z_N(\mathbf{x}) = \sum_k e^{-\lambda_k N}\psi_k(\mathbf{x})c_k(\mathbf{x}'), \quad \hat{g}\psi_k = e^{\varphi-\lambda_k}\psi_k. \qquad (2.9)$$

If the first eigenvalue λ_0 of Eq. (2.9) belongs to a discrete spectrum, i.e., is separated by a finite interval from the next eigenvalue, then for large enough N, the memory of the initial condition is erased, and only one term in the series (2.9) remains:

$$Z_N(\mathbf{x}) \sim e^{-\lambda N}\varphi(\mathbf{x}), \qquad (2.10)$$

$$\hat{g}\psi = e^{\varphi-\lambda}\psi(\mathbf{x}), \qquad (2.11)$$

where λ is the smallest eigenvalue of Eq. (2.11).

In view of the symmetry of the expression $Z\left(\begin{matrix} \mathbf{x}', \mathbf{x} \\ 1, N \end{matrix}\right)$ in $\mathbf{x}$ and $\mathbf{x}'$, we get the following as an asymptotic expression in N:

$$Z\left(\begin{matrix} \mathbf{x}', \mathbf{x} \\ 1, N \end{matrix}\right) = e^{-\lambda N}\psi(\mathbf{x})\psi(\mathbf{x}'). \qquad (2.12)$$

Equation (1.12) implies that $Z_N \sim e^{-\lambda N}$, and hence

$$F = N\lambda = N\lambda\{\varphi\}. \qquad (2.13)$$

If the function $\varphi(\mathbf{x})$ is smooth enough (varying little over distances $\sim a$) and the kernel $g(\mathbf{y})$ is spherically symmetric, then $\hat{g}\psi = \psi + a^2\Delta\psi$, and Eq. (2.12) takes on the form

$$a^2\Delta\psi + (1 - e^{\varphi-\lambda})\psi = 0. \qquad (2.14)$$

If the field vanishes at infinity ($\varphi(\infty) = 0$), then Eq. (2.14) will acquire a discrete eigenvalue when the potential well $\varphi(\mathbf{x}) = u(\mathbf{x})/T$ reaches a critical size. This means that the potential well can retain practically the entire random coil within its own volume, and turn it into a compact globule. If the volume of the well is large enough, then for the first eigenvalue $|\varphi(\mathbf{x}) - \lambda| \ll 1$ (at least, in the region where $\psi(\mathbf{x})$ differs substantially from zero). In this case, (2.14) takes on the form

$$a^2\Delta\psi + (\lambda - \varphi)\psi = 0. \qquad (2.15)$$

Thus, for example, for a deep spherically symmetrical well in which the ground level is near the bottom, where $\varphi(r) = \varphi_0 + \alpha r^2/2$, we obtain

$$\lambda = \varphi_0 + \tfrac{3}{2}a\sqrt{\alpha}, \quad \alpha = (\Delta\varphi)_0 = (\Delta u)_0 / T. \qquad (2.16)$$

At high temperatures, there are no discrete levels in the field $\varphi(\mathbf{x}) = u(\mathbf{x})/T$. As we lower the temperature, a first discrete level $\lambda_c = \lambda\{\varphi\}_c$, $\varphi_c = u(\mathbf{x})/T_c$ will appear, starting at a critical value T_c. In general, this level continuously splits off from the boundary of the continuous spectrum ($\lambda_c = 0$). If the potential well is close to the critical value $\varphi_c(\mathbf{x}) = u(\mathbf{x})/T_c$, then, as we can verify,

$$\lambda\left\{\frac{u}{T}\right\} - \lambda\left\{\frac{u}{T_c}\right\} \sim (T - T_c)^2. \qquad (2.17)$$

Since the well does not hold the random coil when $T > T_c$, and $F_T = 0$, then a second-order phase transition occurs at the point T_c:

$$\frac{F_T}{N} = \lambda T = \begin{cases} A(T - T_c)^2/2, & T < T_c, \\ 0, & T > T_c. \end{cases} \qquad (2.18)$$

Finally, if the random coil is held in a spherical volume $V = (4/3)\pi R^3$ ($R \gg a$) by potential barriers, then

$$\lambda = \frac{\pi^2 a^2}{R^2}, \quad \psi = \frac{\sin kr}{kr}, \quad k = \frac{\pi}{R}. \qquad (2.19)$$

The pressure on the walls holding the random coil here will be[2]

$$p = -\frac{\partial F_T}{\partial V} = -NT\frac{\partial\lambda}{\partial V} = \frac{N\pi a^2 T}{2R^5}, \qquad (2.20)$$

or, upon introducing the volume per link $v = V/N$,

$$\frac{pv}{T} = \frac{2}{3}\left(\frac{\pi a}{R}\right)^2, \quad \left(\frac{a}{R}\right)^2 \gg \frac{1}{N}. \qquad (2.21)$$

Let us return to determining the fundamental thermodynamic characteristics of the random coil. According to (2.4) and (2.13), we have the following for the equilibrium density:

$$\bar{n}(\mathbf{x}) = N\rho(\mathbf{x}), \quad \rho(\mathbf{x}) = \delta\lambda / \delta\varphi. \qquad (2.22)$$

On the other hand, if we write (2.11) in the form

$$(\psi, \hat{g}\psi) = (\psi^2, e^{\varphi-\lambda})$$

---

[2] In order to permit us to proceed from (2.9) to (2.10), the spacing $\delta\lambda = \lambda_1 - \lambda_0$ between the first two eigenvalues of Eq. (2.11) (or (2.15)) had to satisfy the condition $N\delta\lambda \gg 1$. For a spherical potential well of radius R, this gives as the condition of applicability of the derived formulas $(a/R)^2 \gg 1/N$. That is, the condition for the pressure will be $p \gg T/V = p_g/N$, wher p_g is the pressure of a gas at the same values of V, N, and T.

and let it vary over $\delta\psi$, $\delta\varphi$, and $\delta\lambda$, we get

$$(\delta\psi,\ \hat{g}\psi) + (\psi,\ \hat{g}\delta\psi) - 2(\delta\psi,\ \psi e^{\varphi-\lambda}) = (\psi^2 e^{\varphi-\lambda},\ \delta\varphi - \delta\lambda).$$

Taking into account the fact that $(\psi,\ \hat{g}\delta\psi) = (\hat{g}\psi,\ \delta\psi)$, we have

$$\delta\lambda = (\psi^2 e^{\varphi},\ \delta\varphi)\ /\ (\psi^2 e^{\varphi}). \tag{2.23}$$

Hence,

$$\rho(x) \sim \psi^2 e^{\varphi}, \qquad \int \rho(x) d^3x = 1. \tag{2.24}$$

Analogously we obtain for the correlation of density fluctuations

$$\overline{n(x_1)n(x_2)} = -\frac{\delta^2 Z}{\delta\varphi_1\,\delta\varphi_2}\ \Big|\ Z = N^2\frac{\delta\lambda}{\delta\varphi_1}\frac{\delta\lambda}{\delta\varphi_2} - N\frac{\delta^2\lambda}{\delta\varphi_1\,\delta\varphi_2},$$

$$\frac{\overline{n_1 n_2}}{\bar{n}_1 \bar{n}_2} - 1 = -\frac{1}{N}\Big(\frac{\delta^2\lambda}{\delta\varphi_1\,\delta\varphi_2}\ \Big|\ \frac{\delta\lambda}{\delta\varphi_1}\frac{\delta\lambda}{\delta\varphi_2}\Big) \to 0. \tag{2.25}$$

The obtained result shows that when the potential well suffices to hold the random coil, the density $\rho(x) \sim \psi^2 e^{\varphi}$ becomes a thermodynamic characteristic of the coil. Nevertheless, the density $\rho(x)$ does not coincide with the density of the distribution $p_0(x)$ of the endpoints of the random coil. According to (2.12), the latter is

$$p_0(x) = \psi(x)\ \Big/ \int \psi d^3x. \tag{2.26}$$

Thus, for example, according to (2.19), (2.24), and (2.26), for a random coil constrained by potential walls within a spherical volume of radius R, we have

$$\rho(r) \sim \frac{\sin^2 kr}{k^2 r^2}, \qquad p_0(r) \sim \frac{\sin kr}{kr}, \qquad k = \frac{\pi}{R}.$$

In order to trace how the equilibrium density $\rho(x)$ is formed as $N \to \infty$, let us derive Eq. (2.24) by direct examination of the statistical integral. In addition to the expression $Z\begin{pmatrix} x_1, & x_N \\ 1, & N \end{pmatrix}$, which corresponds to the fixed ends of the chain, let us introduce the value $Z\begin{pmatrix} x_1, & x, & x_N \\ 1, & sN, & N \end{pmatrix}$ of the statistical integral for a chain having fixed ends, x being the coordinate of the intermediate link of number sN. From the structure of the statistical integral, we can easily derive the relation

$$Z\begin{pmatrix} x_1, & x, & x_N \\ 1, & sN, & N \end{pmatrix} = Z\begin{pmatrix} x_1, & x \\ 1, & sN \end{pmatrix} e^{\varphi(x)} Z\begin{pmatrix} x, & x_N \\ 1, & (1-s)N \end{pmatrix}. \tag{2.27}$$

If the point sN lies far enough from the ends of the chain, i.e., $sN \gg 1$, and $(1-s)N \gg 1$, then according to (2.12), we get

$$Z\begin{pmatrix} x_1, & x, & x_N \\ 1, & sN, & N \end{pmatrix} \to e^{-\lambda N}\psi(x_1)\psi^2(x)e^{\varphi(x)}\psi(x_N) \sim \psi(x_1)\psi(x_N)\rho(x). \tag{2.28}$$

This means that the probability density $p_{sN}^{(N)}(x)$ coincides with the expression for $\rho(x)$ for links remote enough from the ends of the chain. Since $\rho(x) = \int_0^1 p_{sN}^{(N)}(x)ds$, i.e., it is precisely the links far from the ends that contribute most to the density $\rho(x)$ for a given chain, this explains the obtained result.

3. THERMODYNAMICS OF NON-EQUILIBRIUM STATES OF A RANDOM COIL WITHOUT INTERACTION. VARIATION PRINCIPLE

The results of the last section permit us to construct the thermodynamics of a random coil in a non-equili-

brium state, as given by its density function $n(x) = N\rho(x)$. We can find the entropy of such a state by including an external field $\tilde{\varphi}(x)$ that will give rise to the given equilibrium density $n(x)$. If we multiply (2.11) on the left by ψ and take (2.24) into account, we get[3]

$$\psi\hat{g}\psi = n(x), \qquad \int n d^3x = (n, 1) = N, \tag{3.1}$$

$$\tilde{\varphi} - \tilde{\lambda} = \ln\,(n\ /\ \psi^2) = \ln\,(\hat{g}\psi\ /\ \psi) \quad (\tilde{\varphi} = \varphi_{\{n\}}(x),\ \ \tilde{\lambda} = \lambda\{\tilde{\varphi}\}). \tag{3.2}$$

Equations (3.1) and (3.2) define ψ and $\tilde{\varphi} - \tilde{\lambda}$ in terms of the given density $n(x)$. If the distance a between links (the smoothing radius of ψ by the kernel $g(y)$) is small in comparison with the dimensions R of the random coil, then Eq. (3.1) takes on the form

$$a^2\Delta\psi + \psi - n(x)\ /\ \psi = 0. \tag{3.3}$$

According to (2.6) and (2.13), we get the following expression for the configurational entropy:

$$S = (\tilde{\varphi} - \tilde{\lambda},\ n).$$

Considering n to be an assigned quantity from which ψ and $\tilde{\varphi} - \tilde{\lambda} = \ln\,(\hat{g}\psi/\psi)$ are determined, we obtain the entropy of the non-equilibrium state

$$S\{n\} = (\ln\,(\hat{g}\psi\ /\ \psi),\ n), \qquad \psi\hat{g}\psi = n. \tag{3.4}$$

As we see from (3.4), the entropy density is generally not of a local nature. Such a local nature arises in two limiting cases. For a function $n(x)$ that varies slowly enough over all space,

$$n = \psi\hat{g}\psi \to \psi^2, \qquad \ln\frac{\hat{g}\psi}{\psi} \to a^2\frac{\Delta\psi}{\psi} \ll 1.$$

Hence,

$$S\{n\} \to a^2(\Delta\psi,\ \psi) \to -a^2(\nabla n^{1/2},\ \nabla n^{1/2}). \tag{3.5}$$

In the opposite limiting case where $a \gg R$, in which the smoothing radius is large in comparison with the dimensions of the system and the bonds between the links impose no limitations on the possible configurations, then $\hat{g}\psi = $ const., $\psi \sim n$, and we get the ordinary formula for the entropy of a non-equilibrium ideal gas,

$$S\{n\} = -\int n \ln n d^3x + \text{const.}$$

If we consider the free energy $F = N\tilde{\lambda}$ as a functional of $\tilde{\varphi}$, then according to (2.4) and (2.24) we have

$$\delta F = (n,\ \delta\tilde{\varphi}). \tag{3.6}$$

Thus we get the following for the variation in the entropy $S = (n, \tilde{\varphi}) - F$:

$$\delta S = (\tilde{\varphi},\ \delta n) = (\ln\,(\hat{g}\psi\ /\ \psi),\ \delta n), \qquad (1,\ \delta n) = 0. \tag{3.7}$$

The free energy of the random coil in the external field $\varphi = u/T$ when it has the non-equilibrium density $n(\varphi \neq \tilde{\varphi})$ is

$$F\{\varphi,\ n\} = E\{\varphi,\ n\} - S\{n\}$$
$$= (\varphi,\ n) - S\{n\} = (n,\ \varphi - \ln\,(\hat{g}\psi/\psi)). \tag{3.8}$$

The condition that F should be a minimum as a functional of n for fixed external field, temperature, and

---

[3] The normalization of Ψ remained arbitrary in Eq. (2.11). The choice of normalization $(\chi^2,\ \exp(\Phi - \lambda)) = N$ leads to the non-linear equation (3.1).

number of particles gives

$$(\varphi - \lambda - \ln(\hat{g}\psi/\psi), \delta n) = 0,$$

which leads to the previous equation (2.11). Obviously, we get the same result by finding the extreme value of the entropy $S\{n\}$ for the assigned values of the energy $E\{\varphi, n\}$ and the external field φ.

4. STATISTICAL THERMODYNAMICS OF CHAINS SHOWING A VOLUME INTERACTION

The results of the last section permit us to construct a statistical thermodynamics for chains showing a volume interaction U between the links. We should mention that this is not a case of interaction between adjacent links in the chain, which is already taken into account by fixing the correlations g_j (i.e., the linear memory), but of an additional interaction. It generally occurs between remote regions of the chain that have approached one another because of its bending. For example, the energy U can have the form of a pair interaction depending on the geometric distance between the links that have approached one another, i.e.,

$$U(x_1, \ldots, x_N) = \frac{1}{2} \sum V_{ik}, \quad V_{ik} = V(x_i - x_k). \quad (4.1)$$

In the general case, we shall consider the external field φ also to be included along with the volume interaction U. Thus we find it possible to study also those cases in which the size of the interaction or its sign in the absence of a supplementary field does not permit the formation of a compact globule.

The expression for the free energy can be written in the obvious form:

$$F = -\ln Z, \quad Z = \int e^{-U(\Gamma) - (n,\varphi)} \prod_j g_j d\Gamma. \quad (4.2)$$

In the expression for Z, we can transform from integrating over $d\Gamma$ to functional integration over the smoothed density function $(n = n(x))$. Here we must take into account the fact that the logarithm of the Jacobian of such a transformation is the above-defined entropy $S\{n\}$:

$$Z = \int e^{S\{n\} - (\varphi, n) - U\{n\}} \mathscr{D}n, \quad (4.3)$$

$$S\{n\} = \ln\left(\int \prod_j g_j d\Gamma\right)_{\{n\}},$$

$$e^{-U\{n\}} = \langle e^{-U(\Gamma)}\rangle_{\{n\}} = \left(\int e^{-U(\Gamma)} \prod_j g_j d\Gamma\right)_{\{n\}}\left(\int \prod_j g_j d\Gamma\right)_{\{n\}}^{-1}. \quad (4.4)$$

Since the smoothed density n(x) is now a macroscopic characteristic of the non-equilibrium state, then if we know $U\{n\}$, we can find the free energy by finding the extreme value:

$$F = \min F\{n\}, \quad (n, 1) = N = \text{const}, \quad (4.5)$$

$$F\{n\} = U\{n\} + (\varphi, n) - S\{n\}.$$

Thus, first of all, we should elucidate the meaning and find the value of the averaged quantity $U\{n\}$:

$$U\{n\} = -\ln\langle e^{-U(\Gamma)}\rangle_{\{n\}}. \quad (4.6)$$

Averaging in this way gives an especially simple and graphic result for high enough densities n in two limiting cases.

1. When the forces have a long radius r_0 of interaction $(nr_0^3 \gg 1)$, the energy $U(\Gamma)$ itself is a functional of

the smoothed density n(x), and requires no further averaging:

$$U\{n\} = (U(\Gamma))_{\{n\}} = \frac{1}{2}\int\int n(x)n(x')V(x-x')d^3x\,d^3x' = (\varepsilon(n),1),$$

$$\varepsilon(n) = \frac{1}{2}(n, \hat{V}n) \cong \frac{V_0}{2}n^2, \quad V_0 = \int V(x)d^3x. \quad (4.7)$$

2. When the forces are of small radius $r_0 \ll a$ (in particular, repulsion forces at atomic distances) when the number of particles in the smoothing region is large $(na^3 \gg 1)$.

In this latter case the overwhelming majority of particles in the vicinity of a given particle belongs to distant regions of the chain (see Fig. 1). If we average locally (4.6) over the rapidly fluctuating quantity $U(\Gamma)$, the coefficients $g(y_j)$ remain practically constant, and can be omitted. Hence,

$$\langle e^{-U(\Gamma)}\rangle_{\{n\}} = \left(\int e^{-U(\Gamma)}d\Gamma\right)_{\{n\}}\Big/\left(\int d\Gamma\right)_{\{n\}} = e^{F_0(n,0) - F_0(n,U)}. \quad (4.8)$$

Here $F_0\{n, U\}$ has the meaning of the free energy[4] of a system of links having the same interaction U and density n, but with broken bonds g_j, i.e., having no <u>linear memory</u>:

$$F_0\{n, U\} = (1, f_0(n)) = \int f_0(n)d^3x, \quad (4.9)$$

Here $f_0(n)$ is the free energy per unit volume at the local density n; $F_0\{n, 0\}$ is the free energy of an ideal gas (U = 0) of density n:

$$F_0\{n, 0\} = -S_0\{n\} = -(1, s_0(n)), \quad s_0(n) = -n\ln(n/e). \quad (4.10)$$

Thus, $U\{n\}$ can be represented in a single way in both cited cases:

$$U\{n\} = (1, \varepsilon(n)), \quad \varepsilon(n) = f_0(n) + s_0(n) = f_0(n) + n\ln(e/n). \quad (4.11)$$

If we find the extreme value of $F\{n\}$ under the condition $(1, n) = N$, and take account of the fact that $\partial f_0/\partial n = \mu_0(n)$ is the chemical potential of a particle in a system of local density n, we get an equation for ψ analogous to the previous (2.11):

$$\ln(\hat{g}\psi/\psi) = \Phi(n, x) - \lambda;$$

$$\Phi(n, x) = \varphi(x) + \frac{\partial\varepsilon}{\partial n} = \varphi(x) + \mu_0(n) - \ln n, \quad \psi\hat{g}\psi = n, \quad (4.12)$$

where the external field $\varphi(\hat{x})$ is replaced by the effective self-consistent field $\Phi(n, x)$.

Taking into account the fact that $\ln n = \ln g\psi + \ln\psi$, we can write this same equation in the form

$$\mu_0(n) + \varphi(x) - 2\ln\hat{g}\psi = \lambda = \text{const}. \quad (4.13)$$

When the smoothing radius a is large in comparison with the dimensions of the system $(a \gg R)$ and $g\psi$ = const. (i.e., the memory imposes no limitations on the configuration of the links), the condition (4.13) transforms into the ordinary thermodynamic condition of constancy of the chemical potential in a system having no memory:

$$\mu_0(n) + \varphi(x) = \text{const}.$$

Equation (4.13) is a generalization of this condition

---

[4] Here and below, when speaking of the free energy or chemical potential, we refer to the configurational part of these quantities.

to a system having a linear memory. Together with the relation $n = \psi\hat{g}\psi$, we can consider it as an analog of the equation of state of such a system as a whole.

5. THE RANDOM COIL-GLOBULE PHASE TRANSITION IN A SYSTEM SHOWING A VOLUME INTERACTION

At zero external pressure in the absence of external fields, a loose random coil can collapse into a compact globule owing to volume forces of attraction between the links. However, unless we introduce repulsive forces, the problem of the stable state of the globule under zero external pressure can have no solution, since the system would tend to unlimited contraction. When we take into account the repulsive forces at atomic distances, the dense core of the globule is analogous to a condensed phase. Its fringe, which is subject to zero pressure and is restrained only by the bonds along the chain, is an analog of a gas phase. Thus, such a globule should form a sort of two-phase system.

The concrete form of the volume interaction, which takes into account both the attractive forces at relatively great distances and the repulsive forces at atomic distances, is contained in Eq. (4.12) in the dependence of the chemical potential $\mu_0(n)$ on the density. As we intend to analyze not only the conditions for appearance of a compact globule and its structure, but also the corresponding temperature-dependences, we shall transform in this section from energy characteristics expressed in temperature units to the usual notation, introducing the explicit temperature-dependence. This amounts to replacing

$$F \to \frac{F}{T}, \quad f_0 \to \frac{f_0(n,T)}{T}, \quad \mu_0 \to \frac{\mu_0(n,T)}{T}.$$

The equation describing the state of the entire two-phase, spherically symmetrical globule has the following form in the absence of an external field, according to (4.12):

$$\hat{g}\psi = e^{\Phi(n)-\lambda}\psi, \quad \psi^2 = ne^{\lambda-\Phi(n)},$$
$$\Phi(n) = \mu_0(n,T)/T - \ln n, \quad n = n(r), \quad n(\infty) = \psi(\infty) = 0,$$
$$4\pi \int n(r)r^2 dr = N. \tag{5.1}$$

The latter normalization condition determines the value of λ if a globule exists.

For small n (i.e., for a "gas" phase), the expression for the chemical potential $\mu_0(n, T)$ has the well-known form:

$$\mu_0/T = \ln n + an, \quad a = a(T), \tag{5.2}$$

Here $\alpha(T) > 0$ at high temperatures, and $\alpha(T) < 0$ at low temperatures. In the high-density region, $\mu_0(n)$
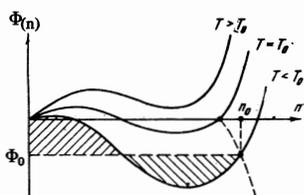

FIG. 2. Schematic graph of the relation $\Phi = \Phi(n)$ for different temperatures. For a globule having a continuous density distribution, $\Phi_0 = \Phi(n_0)$ is found from the condition $\int_0^{n_0} (\Phi - \Phi_0)\, dn = 0$. The temperature T_0 at which appearance of such a globule becomes possible corresponds to the values $\Phi_0 = 0$.

shows a minimum corresponding to a condensed phase. Thus $\Phi(n)$ will be given by

$$\Phi(n) = \begin{cases} an, & an \ll 1, \\ \Phi_m + \gamma(n - n_m)^2 & (n - n_m)/n_m \ll 1. \end{cases} \tag{5.3}$$

Figure 2 shows the characteristic course of $\Phi(n)$.

There is generally a discontinuity in the density $(\delta n = n_2 - n_1)$ in the transition region between the phases, and hence, a discontinuity in $\psi(\delta\psi = \psi_2 - \psi_1)$. We have assumed the smoothing radius a to be large in comparison with the atomic distances. Therefore we can assume, to the same degree of accuracy to which Eq. (4.14) was derived, that $\Phi(n)$ consists of two regions:

$$\Phi(n) = \begin{cases} \mu_{01}(n)/T - \ln n, & n > n_1, \\ \mu_{02}(n)/T - \ln n, & n < n_2, \end{cases}$$

Here ψ and n are continuous within each of them. The exact equation (5.1) for ψ holds everywhere, including the neighborhood of the point of discontinuity (the phase boundary).

As we can convince ourselves, the physical picture for large globules ($R \gg a$) arising from the solution (5.1) has the following form: the core of the globule has practically constant density, reaching some value n_0 at the center, while the transition layer and fringe are of small thickness $\delta R \ll R$ ($R^3 n_0 4\pi/3 \cong N$). Hence, if we assume that $r = R + x$, and take into account the fact that $\hat{g}\psi = \psi + a^2 d^2\psi/dx^2$ on both sides of the narrow boundary layer at the phase boundary, we can write Eq. (5.1) for each of the phases in the form

$$a^2 \frac{d^2\psi}{dx^2} = \psi(e^{\Phi(n)-\lambda} - 1), \quad \psi^2 = ne^{\lambda-\Phi(n)} \tag{5.4}$$

with the boundary conditions:

phase 1: $\quad n(-\infty) = n_0, \quad \psi^2(-\infty) = \psi_0^2 = n_0 e^{\lambda-\Phi_0},$

phase 2: $\quad n(\infty) = 0, \quad \psi^2(\infty) = 0. \tag{5.5}$

Equations (5.4) and (5.5) directly imply that $\lambda = \Phi(n_0) \equiv \Phi_0$, and $\psi_0^2 = n_0$. The first integral of (5.4)–(5.5) has the form

$$a^2\left(\frac{d\psi}{dx}\right)^2 = \begin{cases} H_1^2(\psi) & \text{(I)}, \\ H_2^2(\psi) & \text{(II)}; \end{cases} \tag{5.6}$$

$$H_1^2(\psi) = \int_0^n (\Phi - \Phi_0)dn - q(n), \quad H_2^2(\psi) = \int_{n_3}^n (\Phi - \Phi_0)dn - q(n),$$
$$q(n) = n[e^{\Phi - \Phi_0} - 1 - (\Phi - \Phi_0)], \quad \psi^2 = ne^{\Phi - \Phi_0}. \tag{5.7}$$

Equations (5.6) can be solved by quadratures, and they give

$$a\int_{\psi_1}^{\psi} \frac{d\psi}{H_1(\psi)} = x \quad \text{(I)}; \quad a\int_{\psi_2}^{\psi} \frac{d\psi}{H_2(\psi)} = -x \quad \text{(II)}. \tag{5.8}$$

In particular, far from the phase boundary we obtain an exponential decay:

$$n \sim e^{-\nu_1 x} \quad (x \to \infty), \quad \nu_1 = \frac{1}{a}\sqrt{e^{-\Phi_0} - 1};$$

$$n - n_0 \sim e^{\nu_2 x} \quad (x \to -\infty), \quad \nu_2 = \frac{1}{a}\left[\frac{2n_0\Phi'(n_0)}{1 - n_0\Phi'(n_0)}\right]^{1/2} \tag{5.9}$$

The value of n_0 must be obtained by matching the solutions ψ_1 and ψ_2 at the point $x = 0$. Here we can no

longer use the differential form of Eqs. (5.4), but must start with the exact equation (5.1). As we can convince ourselves, the local chemical potential $\mu_0(n)$ and the local pressure $p = f_0(n) - n\mu_0(n)$ must remain continuous at the boundary in the adopted approximation. Naturally, the following thermodynamic inequality must be satisfied at every point:

$$\frac{1}{T}\frac{\partial\mu_0}{\partial n} = \frac{\partial\Phi}{\partial n} + \frac{1}{n} > 0. \qquad (5.10)$$

Furthermore, the existence of a compact globule must be thermodynamically more favorable than a loose random coil having $n \to 0$ and $F = 0$. For large globules, according to (4.11), this condition is reduced to the elementary requirement:

$$\varepsilon(n_0) = \frac{f_0(n)}{T} - n_0 \ln\frac{n_0}{e} = \int_0^{n_0} \Phi(n)\,dn < 0. \qquad (5.11)$$

First of all, let us find out which cases will show a continuous density distribution throughout the system without a discontinuity at the phase boundary. The right-hand sides of (5.6) differ by the quantity

$$H_1^2 - H_2^2 = \int_0^{n_0} (\Phi - \Phi_0)\,dn.$$

Hence, if there is a solution satisfying (5.5) and continuous throughout space, then n_0 must be fixed by the condition

$$\int_0^{n_0} (\Phi - \Phi_0)\,dn = 0, \qquad (5.12)$$

Together with Eq. (5.11), this gives $\Phi_0 \equiv \Phi(n_0) < 0$. The conditions (5.11) and (5.12) begin to be satisfied when $T \leq T_0$, where T_0 is given by the equations:

$$\Phi(n_0, T_0) = 0, \qquad \int_0^{n_0} \Phi(n, T_0)\,dn = 0.$$

On the other hand, existence of a single solution means that the inequality (5.10) must be satisfied throughout the region $0 < n < n_0$. In other words, the temperature T must be above the critical temperature $(T > T_C)$[5]. Thus, two cases are possible for a large enough chain:

1) $T_0 > T_C$. When $T > T_0$, the equation of state of the two-phase system has no solution, and no globule arises. The globule arises at the point $T = T_0$, with a continuous density distribution, and with an extra free energy vanishing at the transition point, as given by

$$\frac{F}{TV} = \begin{cases} n_0\Phi(n_0), & T < T_0, \\ 0, & T > T_0; \end{cases}$$

$$\int_0^{n_0} (\Phi - \Phi_0)\,dn = 0.$$

This corresponds to a first-order transition. On passing through the temperature T_C, a density discontinuity appears at the boundary between the dense core and the fringe of the globule. This discontinuity vanishes at the critical point T_C. When there is a discontinuity, the condition (5.12) is violated, and this leads to a second-order transition.

2) $T_0 < T_C$. In this case, there is no globule at temperatures $T > T_0$, and on passing through T_0 a globule appears, having a density discontinuity in the transition layer—a first-order transition.

In the case of a continuous density distribution, assuming that $\Phi - \Phi_0 \ll 1$, we can derive from (5.7) and (5.12) the following simple equation for the density throughout the transition region:

$$\frac{a^2}{2n}\left(\frac{dn}{dx}\right)^2 = \int_0^n (\Phi - \Phi_0)\,dn, \qquad \int_0^{n_0} (\Phi - \Phi_0)\,dn = 0,$$

Hence,

$$\frac{x}{a} = \int \left[2n\int_{n_0}^n (\Phi - \Phi_0)\,dn\right]^{-1/2} dn.$$

The discussed model is an extremely simplified system. In particular, substantial changes in the structure of the compact core appear even when we take into account the dependence of the interaction forces and correlations $g(\mathbf{y})$ not only on the relative coordinates, but also on the relative orientation of the links. The latter situation corresponds to increasing stiffness of the chain and possible appearance of orientational ordering in the globule. Many of these problems can be studied analogously.

Even more important changes arise when we take into account the informational memory in a heteropolymer. The existence of such a memory (i.e., a definite sequence of alternation of differing links in the chain) can give rise to spatial linkage of certain quite definite, remote links, which must be considered to be fixed ("an aperiodic crystal"). The fundamental problem arising here consists in the following: do such fixed spatial linkages arise from statistical equilibrium of the system having the given primary chain structure under the given external conditions, or are they only a non-equilibrium or metastable state fixed in greater detail? In the former case, formation of such globules from free chains requires only creation and maintenance of appropriate external conditions. In the latter case, it must arise from a definite complex kinetic mechanism. Finally, the situation may possibly vary as a function of the chain length and the geometric dimensions of the globule. Actually, the existence of strictly fixed spatial linkages of quite definite remote links (not reducible to periodic repetition) implies zero entropy of the system by this criterion. This seems improbable for the equilibrium state of an excessively long chain (owing to the rapid growth in the number of different, almost identically favorable variants of such linkages). If this is so, then critical dimensions should exist for heteropolymeric chains capable of self-organization into a quite definite aperiodic spatial structure (the so-called tertiary structure of the globule). The statistical theory of heteropolymeric chains having volume interactions should answer these questions.