Nonergodic state of heteropolymer globules

A. M. Gutin and E. I. Shakhnovich

Protein Institute, Academy of Sciences of the USSR

(Submitted 26 April 1989)
Zh. Eksp. Teor. Fiz. **96**, 2096–2107 (December 1989)

The replica method is used to study the properties of the globular state of a statistical copolymer consisting of two kinds of units. The properties of such a globule are strongly dependent on the number of spatial dimensions. In spaces with a small number of dimensions $d \leqslant 2$, a nonergodic state with broken replica symmetry is formed, which corresponds to the presence of many minima in the phase space. At these minima, the distance between members of the chain is fixed. Such behavior is observed if different kinds of monomers tend to separate into layers. If the monomers mix, a state with a single energy minimum is formed without the size of the chain link being fixed.

1. INTRODUCTION

The statistical mechanics of disordered heteropolymers is one of the major problems of the theory of disordered systems. The importance of the problem is due to the biological applications to the theory of proteins and nucleic acids. The presence of disorder and of competition of different interactions (frustrations) suggests that the structure of the phase space of such systems can be rather complex — with barriers separating pure states — as observed in spin glasses.[1] Indeed, it was shown in Refs. 2–4 by the replica method that under certain conditions in heteropolymer globules, a transition to a nonergodic state with a disturbed replica symmetry takes place. (The replica method for studying heteropolymer coils was used in Refs. 5 and 6.) It was shown in Refs. 3 and 4 that the properties of the nonergodic phase are strongly dependent on the number of spatial dimensions; these properties were the least trivial for chains in spaces with a small number of dimensions $d < 2$. This is due to the fact that the joining of units into a chain is most appreciable in this case and gives rise to additional frustrations.

It should be noted that in the initial model, used in Refs. 2–4, it was assumed that the number of kinds of monomers was infinite, and the interaction constants of different pairs of monomers were independent random variables. The latter assumption means that instead of specifying the sequence of N quantities (N being the number of monomers in the chain), i.e., a linear memory,[7] it is necessary to specify $N(N-1)/2$ parameters.

The opposite model — that of a disordered heteropolymer composed of two kinds of monomers — was discussed in Ref. 8. The latter studied the behavior of such a heteropolymer in 3-*D* space and showed that this case involves the formation of a domain structure without violation of replica symmetry, i.e., without the formation of a nonergodic state. It is natural to ask the question, will such a state of a "two-letter" heteropolymer exist in a space with a small number of dimensions? It turns out that a nonergodic phase with a large number of equilibrium states arises in this case.

2. MODEL AND BASIC RELATIONSHIPS

Let us consider a polymer chain composed of two kinds of monomers. The statistical sum of such a system is

$$Z = \int \prod_{i=1}^{N} dr_i g(r_i - r_{i-1}) \prod_{i<j} (1 + f_{ij}(r_i - r_j)), \qquad (1)$$

r_i being the spatial coordinates of the ith monomer along the chain; as usual (see, for example, Ref. 7), the function $g(r)$ describes the linking of the monomers into a chain:

$$g(r) = (2\pi a^2)^{-d/2} \exp(-r^2/2a^2), \qquad (2)$$

a being the size of the link between adjacent monomers along the chain; $f_{ij}(r)$ is the Meyer function of interaction between the ith and jth monomers:

$$f_{ij}(r) = \exp(-U_{ij}(r)) - 1,$$

where $U_{ij}(r)$ is the energy of interaction of the ith and jth monomers, and the temperature is set equal to one. The Meyer function in the case of two kinds of monomers may be represented in the form

$$f_{ij}(r) = f_0(r) + (\sigma_i + \sigma_j) f_1(r) + \sigma_i \sigma_j f_2(r), \qquad (3)$$

with $\sigma_i = 1$ if the ith monomer is of the first kind, and $\sigma_i = -1$ in the opposite case.

The free energy of such a chain is a self-averaging quantity. To calculate it, we use the replica method[1]:

$$\langle \ln Z \rangle_{av} = \lim_{n \to 0} \frac{\langle Z^n \rangle_{av} - 1}{n}, \qquad (4)$$

where $\langle \cdots \rangle_{\text{av}}$ denotes averaging over the disorder, i.e., over different sequences of the two kinds of monomers. The averaging on the right-hand side of Eq. (4) can be carried out for an integer-valued n:

$$\langle Z^n \rangle_{av} = \int \prod_{i=1}^{N} \prod_{\alpha=1}^{n} dr_i^\alpha g(r_i^\alpha - r_{i-1}^\alpha) \times \left\langle \prod_{i<j} \prod_{\alpha} (1 + f_{ij}(r_i^\alpha - r_j^\alpha)) \right\rangle_{av}. \qquad (5)$$

Appendix A gives a description of an analogue of a virial expansion by the replica method. In the approximation of broken bonds, which holds when the intrinsic volume of the monomer is $v \ll a^d$, virial expansion gives

$$\langle Z^n \rangle_{av} = \int \prod_{i=1}^{N} \prod_{\alpha=1}^{n} dr_i^\alpha g(r_i^\alpha - r_{i-1}^\alpha) \exp(-E\{Q\}). \qquad (6)$$

Here $E\{Q\}$ denotes a functional having the meaning of energy, of the form

0038-5646/89/121185-07$04.00

$$E\{Q\}=E\{Q_\alpha,\ Q_{\alpha\beta},\ Q_{\alpha\beta\gamma},\dots\},\tag{7}$$

where the following macroscopic parameters (order parameters) are introduced:

$$Q_{\alpha_1\alpha_2\dots\alpha_k}(r_1,r_2,\dots,r_k)$$

$$=\sum_{i=1}^{N}\delta(r_i^{\alpha_1}-r_1)\cdot\dots\cdot\delta(r_i^{\alpha_2}-r_2)\,\delta(r_i^{\alpha_k}-r_k).\tag{8}$$

Obviously, $Q_\alpha(r)$ is the density of the monomers of replica α at the point r, and the parameters with a large number of replica indices are the correlators of the links of the chain of the corresponding replicas. From the definition (8) follows the obvious normalization condition

$$\int Q_{\alpha_1\alpha_2\dots\alpha_k}(r_1,r_2,\dots,r_k)\,dr_1\,dr_2\dots dr_k=N.\tag{9}$$

As already stated, we are interested in the situation where the chain is globularized. In this case, the order parameters Q are slightly fluctuating quantities.[3] Moreover, if the globule is sufficiently large (its volume is $V\gg a^d$), the volume approximation of Ref. 7 can be used. In this approximation, the density $Q_\alpha(r)$ is constant inside the globule and zero outside it. For the order parameters, the following relation holds:

$$Q_{\alpha_1\alpha_2\dots\alpha_k}(r_1,r_2,\dots,r_k)=Q_{\alpha_1\alpha_2\dots\alpha_k}(r_1-r_k,r_2-r_k,\dots).\tag{10}$$

In particular, the two-replica order parameter $Q_{\alpha\beta}(r_1,r_2)$ depends only on the difference between its spatial arguments. It is also important to note that by virtue of the symmetry with respect to the replicas (5), the quantities Q_α are independent of the replica index: $Q_\alpha=\rho$. However, this does not hold for correlators with a large number of replica indices — destruction of the replica symmetry may occur.

Appendix A indicates a method of obtaining the quantity $E\{Q_\alpha,Q_{\alpha\beta},Q_{\alpha\beta\gamma},\dots\}$ in the form of a series in the correlators $Q_\alpha\,Q_{\alpha\beta},Q_{\alpha\beta\gamma}$, etc. We shall see below that for $d<2$, the quantities $Q_{\alpha_1\alpha_2\dots\alpha_k}(r_1-r_k,r_2-r_k,\dots)$ are localized in accordance with spatial arguments with characteristic scale R. Different terms of the series $E\{Q_\alpha\,Q_{\alpha\beta},Q_{\alpha\beta\gamma},\dots\}$ are characterized by different powers of R^{-d}, with the terms corresponding to the ring diagrams having the minimum power (see Appendix B). If the scale R is sufficiently large (a quantitative estimate will be given below), but smaller than $V^{1/d}$, one can readily ascertain that to first order in R^{-d} expression (7) for the energy will be considerably simplified:

$$E\{Q\}=NnE_0(\rho)+E_2\{Q_{\alpha\beta}\},\tag{11}$$

where energy E_2 is determined by the sum of the ring diagrams and hence, is dependent solely on the two-replica order parameters. In analytic form, the energy E_2 becomes

$$E_2=-\sum_{m=2}^{\infty}B_m(\rho)$$

$$\times\sum_{\{\alpha_1,\dots,\alpha_m\}}\int dr_1\dots dr_m Q_{\alpha_1\alpha_2}(r_1-r_2)\,Q_{\alpha_2\alpha_3}(r_2-r_3)$$

$$\dots Q_{\alpha_m\alpha_1}(r_m-r_1).\tag{12}$$

Here the symbol $\{\alpha_1,\alpha_1,\dots,\alpha_m\}$ signifies that the summation is carried out for $\alpha_1\neq\alpha_2,\alpha_2\neq\alpha_3,\dots,\alpha_m\neq\alpha_1$. In the limit of slight heterogeneity (or low density), we can drop all but the first term of the series in m. In this case, we arrive at the model discussed in Ref. 4. It is convenient to go over to momentum space:

$$E_2=-V\sum_{m=2}^{\infty}B_m(\rho)$$

$$\times\sum_{\{\alpha_1,\dots,\alpha_m\}}\int (dp/(2\pi)^d)\,Q_{\alpha_1\alpha_2}(p)\,Q_{\alpha_2\alpha_3}(p)\dots Q_{\alpha_m\alpha_1}(p),\tag{13}$$

where the usual Fourier transformation is used:

$$Q_{\alpha\beta}(p)=\int Q_{\alpha\beta}(r)\exp(ipr)\,dr.$$

It is now easy to rewrite expression (6) in the form

$$\langle Z^n\rangle_{av}=\int\prod_{\alpha\neq\beta}DQ_{\alpha\beta}(r)\exp(-E\{Q\}+S\{Q\}),\tag{14}$$

where $S\{Q\}$ is the entropy corresponding to the number of links of the chain of all the replicas giving the specified $Q_{\alpha\beta}(r)$:

$$S\{Q\}=\ln\int\prod_{i,\alpha}dr_i^\alpha g(r_i^\alpha-r_{i-1}^\alpha)\prod_{\alpha\neq\beta}\delta\Big\{\sum_{i=1}\delta(r_i^\alpha-r_1)\cdot$$

$$\times\,\delta(r_i^\beta-r_2)-Q_{\alpha\beta}(r_1,r_2)\Big\}.$$

Calculation of the entropy defined in this manner is not very difficult and is carried out as in Ref. 7. The difference lies in the fact that instead of an external field, it is necessary to introduce the external potential of the replica interaction $U_{\alpha\beta}(r_1,r_2)$. An entropy calculation is presented in more detail in Ref. 4, which showed that in the approximation in which the ground state dominates[7] we have

$$S\{Q\}=-{}^1\!/_2a^2\sum_\alpha\int(\partial\psi/\partial r_\alpha)^2d\{r\},\tag{15}$$

where $\{r\}$ denotes the set of spatial coordinates of the ends of all n replicas and is a dn-component vector with coordinates $(\{r_1,r_2,\dots,r_n\})$. The two-replica order parameter is easily expressed in terms of the introduced function ψ

$$Q_{\alpha\beta}(r_1,r_2)=\int\psi^2(\{r\})\,\delta(r_\alpha-r_1)\,\delta(r_\beta-r_2)\,d\{r\}.\tag{16}$$

As already noted, in the globular state the order parameters Q are weakly fluctuating, i.e., the average field approximation is valid for them. This approximation corresponds to taking the integral in expression (14) by the method of steepest descents, i.e., to maximizing (not minimizing — see Ref. 1) the free energy

$$F_n\{Q\}=E\{Q\}-S\{Q\}\tag{17}$$

with respect to the parameters Q, or, what is completely equivalent, with respect to the function ψ allowing for Eq. (16) and for the normalization condition (9), which takes the simple form

$$\int\psi^2(\{r\})\,d\{r\}=N.\tag{18}$$

3. QUALITATIVE ANALYSIS

Carrying out the maximization of the functional (17) is a very complicated task. However, much useful information can be extracted from some simple qualitative considerations. We shall first show the qualitatively different behavior of the studied system as a function of the number of space dimensions d. Indeed, consider some function ψ_1 providing a maximum to the functional (17) and satisfying the normalization condition (18). We define the one-parameter family of functions

$$\psi_R(\{r\})=R^{d(1-n)/2}\psi_1(r_1/R, r_2/R, \ldots, r_n/R),$$

which also satisfies the condition (18). As can readily be ascertained, we have

$$F_n\{Q\}=-A_1/R^2+A_2/R^d. \qquad (19)$$

The quantities A_1 and A_2 are independent of R and positive for $n<1$.

It is evident from expression (19) that for $d<2$, the maximum is reached when

$$R=R^*=(2A_1/dA_2)^{1/(2-d)}$$

(naturally, ψ_1 satisfies $R^*=1$). At the same time, for $d>2$, the maximum is reached at $R=0$ or $R=\infty$. It follows that for $d<2$, we have a situation analogous to a second-order phase transition, when there exists a large scale for the order parameters (or for ψ) by virtue of the low degree of polymer heterogeneity. On the other hand, for $d>2$, there is an abrupt transition in scale for a finite value of heterogeneity. Only the case $d<2$ will be considered in the remaining part of this work.

It will be shown that for $d<2$, the functional (17) reaches a maximum on the function ψ nonsymmetric in the replica variables. Indeed, assuming the opposite, one can easily understand from the translation invariance of Eq. (10) that the second, entropic term in Eq. (17) will be of the order of unity when $n\to 0$, whereas the first, energy term will be of the order of n. This means that on the symmetric functions ψ the entropy losses are higher than the energy gain. Hence, the maximizing functions ψ is nonsymmetric. According to the standard interpretation,[1] destruction of replica symmetry signifies the existence of a large number of equilibrium states in the system under consideration.

4. DESTRUCTION OF REPLICA SYMMETRY

Since the minimizing function ψ cannot be symmetric in the replica variables, as the zero approximation it is necessary to use Parisi's first stage of destruction of replica symmetry.[8] In this approximation, n replicas are broken up into n/x groups with x replicas in each group $(0<x<1)$, so that the maximization is carried out on test functions of the form

$$\psi(r_1,\ldots,r_n)=\prod_{k=1}^{n/x}\psi_0(r_1^k,\ldots,r_x^k),$$

where the coordinate r_1^k corresponds to the l th replica from the k th group, and the function ψ_0 is symmetric in its arguments.

It is easy to understand that in such an approximation, the order parameter $Q_{\alpha\beta}$ is zero (in the thermodynamic limit, when $V\gg R^d$) if the replicas α and β are different groups, whereas for replicas from the same group $Q_{\alpha\beta}(r)$ is independent of either the choice of the replicas or the choice of the group. Hence it follows immediately that the free energy (17) of the replicas is

$$F_n\{Q\}=(n/x)F_x\{Q\}, \qquad (20)$$

where F_x is the free energy of x replicas, also determined by expressions (11)–(17), where instead of n it is necessary to use x.

Before turning to further qualitative analysis, it is essential to determine the coefficients B_m in the expression (12) for the energy. For simplicity we assume that first, the probabilities of occurrence of the two kinds of monomers are equal, and second, in expression (3) for the Meyer functions we have $f_1=0$. The latter fact means that the interaction of monomers of one kind is the same as that of the other kind. In Appendix B it is shown for such a case that

$$B_m=B^m/2m.$$

To clarify the meaning of the constant B in this expression, let us consider a mixture of monomers forming a chain, but without polymer linkages. It is simply a mixture of two kinds of molecules. The free energy per unit volume in this case is a function $f(\rho_1,\rho_2)$ of the densities of both components. It is convenient to define the following quantity[7]:

$$f^*(\rho_1,\rho_2)=f(\rho_1,\rho_2)-f_{id}(\rho_1)-f_{id}(\rho_2),$$

where $f_{id}(\rho_i)=\rho_i\ln(\rho_i/e)$ is the free energy of an ideal gas of density ρ_i. One can treat f^* as a function of the total density $\rho=\rho_1+\rho_2$ and of the densities difference $\nu=\rho_1-\rho_2$. The result obtained in Appendix B may be expressed as

$$B=-(\partial^2 f^*/\partial\nu^2)_{\nu=0}. \qquad (21)$$

Thus the physical meaning of B is very simple: for $B>0$, the system tends to separate into phases consisting of like monomers; for $B<0$, it tends to mix. It should be noted that in a real mixture of two components, the separation instability takes place when

$$(\partial^2 f/\partial\nu^2)_{\nu=0}<0,$$

i.e., for $B\rho>1$. It is now easy to carry out the summation in the expression for the energy (13)

$$E_2=-{}^1/_2V\int_0^B du\int(2\pi)^{-d}dp\sum_{\alpha\neq\beta}Q_{\alpha\beta}(p)G_{\alpha\beta}(p), \qquad (22)$$

where the quantities $G_{\alpha\beta}$ satisfy the simple equations

$$G_{\alpha\beta}=u\Big(Q_{\alpha\beta}+\sum_\gamma{}'Q_{\alpha\gamma}G_{\gamma\beta}\Big),\quad \alpha\neq\beta,$$
$$G_{\alpha\alpha}=u\sum_\gamma{}'Q_{\alpha\gamma}G_{\gamma\alpha}, \qquad (23)$$

and the prime at the summation sign denotes summation over $\gamma\neq\alpha$. As already stated, within the same group of replicas the parameters $Q_{\alpha\beta}$ are independent of the replica indices: $Q_{\alpha\beta}=Q$. It is evident that a similar statement applies to the quantities $G_{\alpha\beta}$: $G_{\alpha\beta}=G$, $\alpha\neq\beta$, $G_{\alpha\alpha}=G_0$. Taking this into account, we can substantially simplify Eqs. (23):

$$G=uQ(1+(x-2)G+G_0),$$
$$G_0=u(x-1)QG. \qquad (23a)$$

Solving the system obtained, we get

$$G=uQ[(1+uQ)(1+(1-x)uQ)]^{-1},$$
$$G_0=-u^2Q^2(1-x)[(1+uQ)(1+(1-x)uQ)]^{-1}.$$

Finally, substituting this result into Eq. (22), we obtain

$$E_2=-{}^1/_2V\int_0^B du\int (2\pi)^{-d}\,dp\,(1-x)uQ(p)$$
$$\times[(1+uQ(p))(1+(1-x)uQ(p))]^{-1}. \qquad (24)$$

After these simplifications, the problem of maximizing the free energy is still complicated. It seems natural, therefore, to try to carry out the maximization on test Gauss functions

$$\psi(r_1,\dots,r_x)=A\exp\Big(\sum_{\alpha,\beta}k_{\alpha\beta}r_\alpha r_\beta\Big), \qquad (25)$$

where A is the normalization constant. From the translation invariance of (10) follows the relation

$$\sum_{\alpha=1}^{x}k_{\alpha\beta}=0.$$

Taking into consideration the symmetry of ψ in x replicas, we readily obtain

$$k_{\alpha\alpha}=k, \qquad k_{\alpha\beta}=k/(1-x), \qquad \alpha\neq\beta.$$

From the normalization condition (18), it is now easy to determine A:

$$A=\rho[kx/(1-x)]^{(1-x)d/2}.$$

We substitute the test function (25) into the expressions (16) for $Q_{\alpha\beta}$ and (15) for the entropy and carry out the simple Gaussian integrals, thus obtaining

$$Q(r)=\rho(2\pi)^{-d/2}R^{-d}\exp(-r^2/2R^2),$$
$$Q(p)=\rho\exp(-R^2p^2/2), \qquad (26)$$
$$S_x={}^1/_2Na^2d(1-x)/R^2,$$

where $R^2=(x^{-1}-1)/k$. Finally, we substitute the expressions obtained for $Q(p)$ into Eq. (24), introduce the new integration variables $q=Rp$ and $z=\rho u$, and thus obtain the final expression for the free energy:

$$F_n=nN(E_0+e/R^d-s/R^2), \qquad (27)$$

where

$$e=\frac{1}{2(2\pi)^d\rho}$$
$$\times\int_0^{B\rho}z^{-1}\,dz\int dq\,\frac{z\exp(-q^2/2)}{1+z\exp(-q^2/2)}\,\frac{(1-x)z\exp(-q^2/2)}{1+(1-x)z\exp(-q^2/2)}$$
$$s=a^2d(1-x)/2x.$$

The maximum free energy over R is reached for

$$R=R^*=(2s/de)^{1/(2-d)};$$

substituting this result into Eq. (27), we reduce the problem to the maximization over x of the expression

$$x^{d/2}(1-x)^{1-d/2}$$
$$\times\int_0^{B\rho}dz\int dq\,\frac{z\exp(-q^2)}{(1+z\exp(-q^2/2))(1+(1-x)z\exp(-q^2/2))}.$$

This can easily be done in three limiting cases:

1 – $B\rho\ll1, \quad x\approx d/2,$
2 – $B\rho\gg1, \quad x\approx1-C_1(B\rho)^{-1},$
3 – $-1\leqslant B\rho\leqslant0, \quad 1+B\rho\ll1, \quad x\approx C_2(1+B\rho).$

where C_1 and C_2 are constants of the order of unity, dependent solely on the number of spatial dimensions d. In the limiting cases, the scale R is also easy to determine:

1 – $R\sim(a^2/\rho B^2)^{1/(2-d)},$
2 – $R\sim(a^2/B)^{1/(2-d)},$
3 – $R\sim[a^2\rho(1+B\rho)^{-d/2}]^{1/(2-d)}.$

It is convenient to introduce the dimensionless parameter $\varepsilon=N_R/N_g$, where $N_R=\rho R^d$ is the number of monomers inside a volume of size R, and $N_g=(R/a)^2$ is the number of monomers in an ideal coil of size R. This parameter can serve as a measure of the localization of the monomers. In the limiting cases discussed we have the following asymptotic forms for ε: 1—$\varepsilon\sim(B\rho)^2$, 2—$\varepsilon\sim B\rho$, 3—$\varepsilon\sim(1+B\rho)^{d/2}$. Figure 1 is a qualitative representation of the localization measure ε as a function of the dimensionless parameter $B\rho$, which characterizes the degree of heterogeneity of the chain.

Let us examine certain approximations used in obtaining the results described. First of all, it is necessary to estimate the contribution of the discarded terms in going from Eq. (7) to Eq. (11). It can readily be shown that the discarded terms are small for $B/R^d\ll1$. Note that the average-field approximation is valid for $\rho a^d\gg1$ (Ref. 7), so that in the first and third cases the condition that the discarded terms be small is automatically satisfied. In the second limiting case, the smallness of the discarded terms is ensured by the condition $B\rho\ll\rho a^d$. It is also easy to ascertain that for $B\rho\ll\rho a^d$, the relation $R\gg a$ holds in all limiting cases. This means that the results obtained are independent of the specific form of the function $g(r)$ (2).

5. DISCUSSION

It is necessary first of all to note the qualitatively different behavior of the studied system as a function of the number of spatial dimensions d. These differences are due to the obvious fact that for $d<2$, monomers which are close along

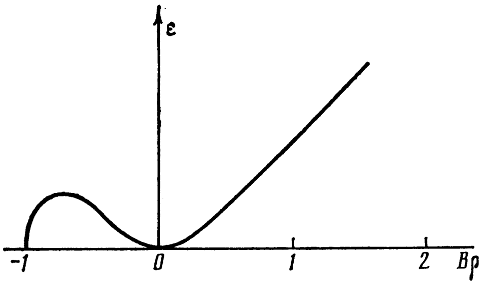


FIG. 1. Localization parameter as a function of separation tendency.

the chain are also close in space, whereas for $d > 2$ the majority of contacts take place between monomers widely separated along the chain. Therefore for $d < 2$ we encounter strong polymeric frustrations, i.e., competition between paired short-range interactions of the monomers and the linking of the monomers into a chain. Obviously, frustrations of this kind do not arise for $d > 2$. It is also important to note that topological limitations in the globularization kinetics of the polymer chain in real 3-D space lead to the formation of a long-lived "folded globule" state, in which a majority of the contacts correspond to monomers that are close to each other along the chain.[9] It seems natural, therefore, that the results which we obtained for $d < 2$ are qualitatively applicable to this case as well.

According to the standard interpretation,[10] the destruction of replica symmetry in the case $d < 2$ signifies the existence, in the system under consideration, of a large number of equilibrium states separated by energy barriers tending to infinity in the thermodynamic limit. The number of states is greater, the larger the parameter x. In particular, at $x = 0$ (when there is no destruction of replica symmetry), we have only one equilibrium state. In our case, the states are characterized by a canonical (frozen) spatial scale of the chain of order R, i.e, in each state the position of the monomer in space fluctuates by an amount on the order of R.

In the framework of such an interpretation, the results obtained in the preceding section are easy to understand. For $B\rho = 0$, i.e., when we are dealing with a homopolymer, $R = \infty$ holds, i.e., there is no canonical scale for the chain in space. In the case of a low degree of chain heterogeneity ($|B\rho| \ll 1$) we find a slight fixation of the scale of the chain in space ($R \to \infty$ as $|B\rho| \to 0$). For $B\rho > 0$, separation of the monomers into phases is energetically favored; this is hindered by the linking of the monomers into a chain. As the separation tendency increases ($B\rho$ increases), the energy barriers also are raised, leading to an increase in the number of states ($x \to 1$ as $B\rho \to \infty$) and in the degree of their freezing ($R \to 0$ as $B\rho \to \infty$). Subtler effects are present when mixing is advantageous ($B\rho < 0$). In a random sequence of monomers along the chain, regions enriched with like monomers occur. For $B\rho = 0$, this is responsible for the fact that in the spatial structure of the globule as well, regions arise that are enriched with like monomers. For $B\rho < 0$, some of these regions begin to mix, and this again is hindered by the linking of the monomers into a chain. As $B\rho$ increases, more and more regions mix, resulting in a decrease in the number of states and degree of their frozenness. Finally, at $B\rho = -1$, all the regions are found to be mixed, and there exists a single state ($x = 0$) without a reliable chain course ($R = \infty$). It is natural to postulate that also for $B\rho < -1$, we have a single state without a definite chain separation scale.

APPENDIX A

VIRIAL EXPANSION IN THE REPLICA METHOD

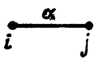
We consider the part of the integrand in Eq. (3) that corresponds to the interaction of replicas:

$$z = \left\langle \prod_{i<j} \prod_{\alpha} (1 + f_{ij}^{\alpha}) \right\rangle_{av}, \tag{A1}$$

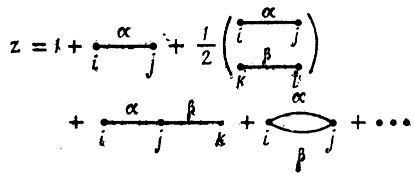
where for brevity the notation $f_{ij}^{\alpha} = f_{ij}(r_i^{\alpha} - r_j^{\alpha})$ has been introduced. Carrying out the multiplication on the right-hand side of (A1), we obtain the series

$$z = 1 + \sum_{\alpha} \sum_{i<j} \langle f_{ij}^{\alpha} \rangle_{av} + {}^1/_2 \sum_{\alpha,\beta} \sum_{\substack{i<j \\ k<l}} \langle f_{ij}^{\alpha} f_{kl}^{\beta} \rangle_{av} + \ldots. \tag{A2}$$

It is convenient to introduce a graphical notation of the terms of the series by correlating the multiplier f_{ij}^{α} with a segment of type α with ends i and j:



In this notation, the series (A2) takes the form



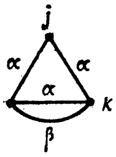
(A3)

here summation over the indices which enter into the diagram and averaging over the disorder is implied.

Virial expansion makes sense when the size of a polymer linkage is much greater than the characteristic scale of the Meyer function $f_{ij}(r)$. In this case, instead of Meyer functions, use may be made of δ functions with the corresponding numerical multipliers. For this purpose, an arbitrary diagram from (A3) is decomposed into connected subdiagrams containing only one type of line (corresponding to a single replica). Each such subdiagram contains indices corresponding to the numbers of the monomers: $i_1, i_2, \ldots, i_m$, where m is the number of vertices in the subdiagram. We then put the subdiagram with α lines in correspondence with the following product of the δ function:

$$\delta(r_{i_1}^{\alpha} - r_{i_2}^{\alpha})\,\delta(r_{i_1}^{\alpha} - r_{i_3}^{\alpha})\ldots\delta(r_{i_1}^{\alpha} - r_{i_m}^{\alpha}), \tag{A4}$$

and we put the entire diagram in correspondence with the product of the expressions, (A4), which correspond to all the subdiagrams. For example, to the diagram



(A5)

corresponds the following product of δ functions:

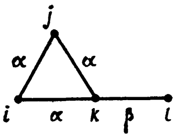
$$\delta(r_i^{\alpha} - r_j^{\alpha})\,\delta(r_i^{\alpha} - r_k^{\alpha})\,\delta(r_i^{\beta} - r_k^{\beta}).$$

To calculate the numerical multiplier corresponding to the diagram, it is necessary to integrate the analytic expression corresponding to each subdiagram with respect to the variables $r_{i_1}^{\alpha}, \ldots, r_{i_m}^{\alpha}$. We then multiply the integrals obtained for all the subdiagrams and average over the disorder. The resulting numerical multiplier is obviously independent of the indices entering into the diagram. For the diagram (A5) this multiplier is

$$p = \langle f_{ij}(r_1 - r_2) f_{jk}(r_2 - r_3) f_{ik}(r_1 - r_3) f_{ik}(r_4 - r_5)\, dr_1 dr_2 dr_4 \rangle_{av}.$$

Finally, the rules for calculating the combinatorial coefficient do not differ from the standard ones.[11]

The next step in obtaining the virial expansion consists in taking the logarithm of the series obtained. As usual, this will lead to a reduction of the unconnected diagrams. In addition, there will be a change in the numerical multipliers in diagrams consisting of more than one type of line, i.e., diagrams corresponding to the interaction of replicas. To describe these changes, we shall consider an arbitrary diagram of this type. Lines of type $\alpha,\beta,\ldots,\gamma$ issue from an arbitrary vertex of the diagram. We divide these types arbitrarily into two groups. We shall define the operation of reduction of the diagram, by requiring that in place of the initial diagram, we obtain a new diagram which differs from the initial one only in having two arbitrary vertices instead of one; all lines of the first-group type issue from one vertex, and all the lines of the second-group type issue from the second vertex. For example, reduction of diagram (A5) results in the diagram



(A6)

Applying the reduction operation to new diagrams, we ultimately arrive at the final unconnected diagram, whose connected parts are the subdiagrams discussed above. We now consider the initial diagram and all the diagrams that can be obtained from the initial one by successively applying several reduction operations to it. We say that the diagram belongs to the *i*th level if it is obtained from the initial one after *i* reduction operations are performed. We subtract from the multiplier of all the diagrams the multiplier of the final diagram (belonging to the last level). We consider an arbitrary diagram belonging to the next-to-the-last level. We subtract its multiplier from the multipliers of all the diagrams which yield the given one after several reduction operations. We do the same with all the diagrams of the next-to-the-last level. We repeat this procedure for all the levels, going up from one level to the next. We finally subtract from the multiplier of the initial diagram the multipliers of all the diagrams obtained from the initial one after a single reduction operation. The difference obtained will be the desired numerical multiplier of the initial diagram after the logarithm of the series has been taken.

Two remarks are in order. First, if we are dealing with a homopolymer, then as a result of the procedure described, the diagrams corresponding to the interaction of replicas will have zero multipliers. Second, if only one diagram can be obtained from the initial diagram as a result of a single reduction operation (this holds for ring diagrams, discussed in Appendix B), the resulting multiplier is equal to the difference of the multipliers of these two diagrams.

We define the macroscopic parameters

$$Q_{\alpha_1\alpha_2\ldots\alpha_k}(r_1, r_2, \ldots, r_k) = \sum_{i=1}^{N} \delta(r_i^{\alpha_1}-r_1)\,\delta(r_i^{\alpha_2}-r_2)\cdot\ldots\cdot\delta(r_i^{\alpha_k}-r_k).$$

It is now easy to obtain the virial expansion in the powers of these parameters. For this purpose, it is necessary to change from summation in the diagrams over $i<j$ to summation over $i\neq j$, introducing the additional multiplier 1/2 for each such pair of indices. It is then necessary to extend the summation also to $i=j$, by introducing an error which is small in the thermodynamic limit. Thereafter, it is easy to transform the sum over the indices corresponding to the numbers of the monomers of the product of the δ functions, corresponding to an arbitrary diagram, into an integral of the product of the macroparameters introduced. Thus, for example, the integral

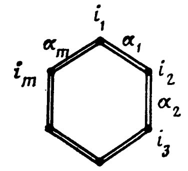
$$\int Q_{\alpha\beta}^2(r_1, r_2)\, Q_\alpha(r_1)\, dr_1\, dr_2.$$

will correspond to the diagram (A5). We thus obtain the desired virial expansion in powers of the macroparameters.

APPENDIX B

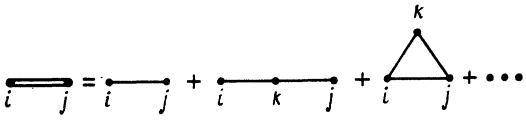
CALCULATION OF THE COEFFICIENTS B_m

It is easy to understand from Appendix A that ring diagrams of the form



(B1)

correspond to the terms of the virial expansion (12). The double line denotes the sum of all the diagrams composed of the same type of lines, with two fixed indices:



(B2)

Here again, summation over all indices except *i* and *j* and integration with respect to the corresponding coordinates is. In analytic form the equality (B2) is

$$F_{ij}(r_i-r_j) = f_{ij}(r_i-r_j) + {}^1/_2 \sum_k f_{ik}(r_i-r_k)\, f_{kj}(r_k-r_j)\, dr_k + {}^1/_2 \sum_k f_{ij}(r_i-r_j)\, f_{ik}(r_i-r_k)\, f_{kj}(r_k-r_j)\, dr_k + \ldots. \qquad \text{(B3)}$$

According to the rules described in Appendix A, to diagram (B1) in the virial expansion correspond the combinatorial multiplier $(2m)^{-1}$ and numerical multiplier B^m, where

$$B = \langle \sigma_i \sigma_j F_{ij}(r)\, dr \rangle_{av}. \qquad \text{(B4)}$$

To calculate B we consider the following statistical sum:

$$Z(h) = \sum_{\{\sigma_i\}} \prod_{i=1}^{N} dr_i \exp\Big(h \sum_{i=1}^{N} \sigma_i\Big) \prod_{i<j} (1 + f_{ij}(r_i-r_j)), \qquad \text{(B5)}$$

where $f_{ij} = f_0 + f_2 \sigma_i \sigma_j$ [see Eq. (3)], the summation in Eq. (B5) is carried out over all $\sigma_i = \pm 1$, and the integration is carried out over the volume V. It is easy to ascertain directly, by expanding Eq. (B5) in a power series of f_{ij}, that

$$B = VN^{-1}(N^{-1}\partial^2 \ln Z(h)/\partial h^2 - 1). \qquad \text{(B6)}$$

On the other hand, in the thermodynamic limit the statistical sum (B5) is identical to the statistical sum of a mixture of two kinds of molecules (details are given in the main text) with densities

$$\rho_1=\rho(1+\exp(-h))^{-1}, \quad \rho_2=\rho(1+\exp(h))^{-1},$$

were $\rho = N/V$. The derivation of Eq. (21) from Eq. (B6) after this remark presents no difficulties.

[1] K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).
[2] T. Garel and H. Orland, Europhys. Lett. **6**, 307 (1988).
[3] E. I. Shakhnovich and A. M. Gutin, Europhys. Lett. **8**, 327 (1989).
[4] E. I. Shakhnovich and A. M. Gutin, J. Phys. A **21**, 1647 (1989).
[5] S. P. Obukhov, J. Phys. A **19**, 3655 (1986).
[6] A. Fernandez and H. Rabitz, Biophys. Chem. **28**, 89 (1987).
[7] I. M. Lifshitz, A. Yu. Grosberg, and A. R. Khokhlov, Rev. Mod. Phys. **50**, 683 (1979).
[8] E. I. Shakhnovich and A. M. Gutin, J. Phys. (Paris) **50**, 1843 (1989).
[9] A. Yu. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, J. Phys. (Paris) **49**, 2095 (1988).
[10] M. G. Mezard, G. Parisi, N. Sourlas *et al.*, J. Phys. (Paris) **45**, 843 (1984).
[11] R. Balescu, *Nonequilibrium Statistical Mechanics*, Wiley-Interscience, New York (1962).

Translated by Adam Peiperl