Frozen states of a disordered globular heteropolymer

E I Shakhnovich and A M Gutin

Institute of Protein Research, USSR Academy of Sciences, Puschino, Moscow Region, USSR

Abstract. The disordered heteropolymer chain with a random sequence of links—a simple model of globular protein—is investigated using the replica approach in the mean-field approximation. It is shown that for low dimensions of space ($d < 2$) polymer bonds play the main role of creating frustrations. The mean-field equations are derived; the solution of these equations is obtained by the variational principle. This solution demonstrates the replica-symmetry breaking with unusual behaviour of the Parisi order parameter $q(x) \sim x^m$ for $x < x_0$ and $m > 1$. It follows that the set of equilibrium states is ultrametric with an anomalously large number of weakly overlapping states.

1. Introduction

The main biological macromolecules—proteins and DNA—are non-uniform linear chains. Proteins in living cells possess a unique three-dimensional structure which is determined by their primary structure: the sequence of their monomers. Revealing this 'physical code', i.e. understanding how the primary structure encodes for the fold of a protein chain is one of the most intriguing unsolved problems of molecular biology (Anfinsen 1973, Creighton 1985). It is widely accepted (see, e.g., Paine and Scheraga 1985, Rammal *et al* 1986 and references therein) that the protein folding problem is analogous to the optimisation problem for the search of the lowest kinetically available minimum of energy.

Hence, it seems interesting to investigate the conformational space of a protein chain, i.e. to determine the number of energy minima and the possible barriers between them.

From a general point of view, the protein can be treated as a frustrated disordered system. Frustrations appear for two reasons: first, interactions between links which are adjacent in space depend on their types. This causes the appearance of frustrated loops in many-particle clusters. The second reason is the influence of polymeric bonds on the probability of contact between groups, i.e. competition between short- and long-ranged interactions (to avoid misunderstanding, we note here that the interactions between groups which are adjacent *along the chain* are called short-ranged interactions).

Significant progress has been achieved in studying disordered frustrated systems such as spin glasses (see the review of Binder and Young (1986)). It seems reasonable to apply the ideas accepted in spin-glass theory to the investigation of non-uniform polymeric chains. This was done by Bryngelson and Wolynes (1987). However, the discussion in that work was based on a phenomenological model analogous to the random energy model (REM) (Derrida 1980). Our aim is to investigate a microscopic model of a linear heteropolymer using the replica approach. Most attention must be

paid to the collapsed (globular) state of a heteropolymer chain. (The replica approach was applied by Fernandez and Rabitz (1987) for investigation of the heteropolymeric coil and by Obukhov (1986) for studying the coil–globule transition.)

Simple estimations which are given below show that the results depend drastically on the space dimensionality d. This is due to the fact that for $d < 2$ the groups adjacent along the chain become adjacent in space also. This is not the case for $d > 2$. Thus the strongest influence of polymeric bonds is for the case when $d < 2$. Here we shall pay most attention to this case.

2. The model and basic relations

We shall use the conformational space approach to describe the heteropolymer chain (Edwards 1966, Kholodenko and Freed 1983). The partition function of such a chain can be written in the continual form

$$Z = \int_{x(0)=0} \exp(-H\{x(\tau)\}) \, dx(\tau) \tag{1}$$

where integration in (1) is taken over all configurations of the chain $x(\tau)$; the conventional Hamiltonian $H\{x(\tau)\}$ (see e.g. Kholodenko and Freed 1983)

$$H = \tfrac{1}{2}a^{-2} \int_0^N (dx(\tau)/d\tau)^2 \, d\tau + \tfrac{1}{2} \int_0^N B(\tau, \tau') \delta(x(\tau) - x(\tau')) \, d\tau \, d\tau'$$

$$+ \tfrac{1}{6}C \int_0^N \delta(x(\tau) - x(\tau')) \delta(x(\tau) - x(\tau'')) \, d\tau \, d\tau' \, d\tau'' \tag{2}$$

is used, where N is the contour length of a chain, a is the effective size of a link, $B(\tau, \tau')$ is the two-body interaction constant between the links positioned in τ and τ' along the chain and C is the three-body interaction constant which is assumed to be independent of the type of the links. The δ function reflects, as usual, the short-ranged nature of the interaction potential; it must be smoothed properly inside the microscopic scales less than $v^{1/d}$ where v is the excluded volume of a link.

The simplest but non-trivial way to introduce the disorder is to assume $B(\tau, \tau')$ to be independent random variables with Gaussian distribution:

$$P(B(\tau, \tau')) = (2\pi B^2)^{-1/2} \exp[-(B(\tau, \tau') - B_0)^2/2B^2] \tag{3}$$

where, as usual, B_0 is the mean and B is the standard variance.

Now we proceed to the evaluation of the free energy using the replica trick:

$$\langle \ln Z \rangle_{av} = \lim_{n \to 0} \frac{\langle Z^n \rangle_{av} - 1}{n}. \tag{4}$$

Here $\langle \ \rangle_{av}$ denotes averaging over disorder with the weight (3).

The averaging in the RHS of (4) is straightforward for integer n; it gives

$$\langle Z^n \rangle_{av} = \int \exp(-H_{eff}\{x_\alpha(\tau)\}) \prod_\alpha Dx_\alpha(\tau). \tag{5}$$

Here α is the replica index and $H_{eff}\{x_\alpha(\tau)\}$ is the effective Hamiltonian which depends on the configurations $\{x_\alpha(\tau)\}$ of all n replicas. It can be presented in the following form:

$$H_{eff} = H_0 + H_1 + H_2$$

where

$$H_0 = \tfrac{1}{2} a^{-2} \int_0^N \sum_\alpha (\mathrm{d}x_\alpha(\tau)/\mathrm{d}\tau)^2 \, \mathrm{d}\tau \qquad (6)$$

is the elastic energy of n replicas,

$$H_1 = \tfrac{1}{2}\tilde{B} \int_0^N \sum_\alpha \delta(x_\alpha(\tau) - x_\alpha(\tau')) \, \mathrm{d}\tau \, \mathrm{d}\tau'$$

$$+ \tfrac{1}{6} C \int_0^N \sum_\alpha \delta(x_\alpha(\tau) - x_\alpha(\tau')) \delta(x_\alpha(\tau) - x_\alpha(\tau'')) \, \mathrm{d}\tau \, \mathrm{d}\tau' \, \mathrm{d}\tau'' \qquad (7)$$

is the one-replica part of the effective Hamiltonian, and $\tilde{B} = B_0 - \tfrac{1}{2}B^2 v^{-d}$.

$$H_2 = -\tfrac{1}{2}B^2 \int_0^N \sum_{\alpha \neq \beta} \delta(x_\alpha(\tau) - x_\alpha(\tau')) \delta(x_\beta(\tau) - x_\beta(\tau')) \, \mathrm{d}\tau \, \mathrm{d}\tau' \qquad (8)$$

is the two-replica part of the effective Hamiltonian; it describes the interaction between replicas; all the non-trivial effects occur due to this term.

We define now the macroscopic (order) parameter:

$$Q_{\alpha\beta}(r_1, r_2) = \int_0^N \delta(x_\alpha(\tau) - r_1)\delta(x_\beta(\tau) - r_2) \, \mathrm{d}\tau. \qquad (9)$$

The order parameter $Q_{\alpha\beta}$ is the correlator for chain passways of replicas α and β. It is more complicated than the overlap between replicas used in spin-glass theory (Sherrington and Kirkpatrick 1975) since it is the matrix of functions of two space coordinates. It is convenient to define the overlap between replicas α and β:

$$q_{\alpha\beta} = v^d N^{-1} \int_0^N \delta(x_\alpha(\tau) - x_\beta(\tau)) \, \mathrm{d}\tau = v^d N^{-1} \int Q_{\alpha\beta}(r, r) \, \mathrm{d}r. \qquad (10)$$

Equation (5) can now be rewritten in terms of $Q_{\alpha\beta}$:

$$\langle Z^n \rangle_{av} = \int \exp(-E\{Q_{\alpha\beta}\} + S\{Q_{\alpha\beta}\}) \prod_{\alpha \neq \beta} DQ_{\alpha\beta}. \qquad (11)$$

The effective energy E in (11) is given by

$$E\{Q_{\alpha\beta}\} = \frac{\tilde{B}}{2} \int \sum_\alpha \rho_\alpha^2(r) \, \mathrm{d}r + \frac{C}{6} \int \sum_\alpha \rho_\alpha^3(r) \, \mathrm{d}r + \frac{B^2}{2} \int \sum_{\alpha \neq \beta} Q_{\alpha\beta}^2(r_1, r_2) \, \mathrm{d}r_1 \, \mathrm{d}r_2$$

where $\rho_\alpha(r)$ is usually defined as

$$\rho_\alpha(r) = \int_0^N \delta(x_\alpha(\tau) - r) \, \mathrm{d}\tau$$

being the density of replica α, and

$$S\{Q_{\alpha\beta}\} = \ln \int \prod_{\alpha \neq \beta} \delta\left(\int \delta(x_\alpha(\tau) - r_1)\delta(x_\beta(\tau) - r_2) \, \mathrm{d}\tau - Q_{\alpha\beta}(r_1, r_2) \right)$$

$$\times \exp(-H_0\{x_\alpha\}) \prod_\alpha Dx_\alpha(\tau) \qquad (12)$$

is the entropy corresponding to the number of chain passways of all replicas which lead to the given set $\{Q_{\alpha\beta}\}$.

The order parameter $Q_{\alpha\beta}(r_1, r_2)$ is subjected to the conditions which follow from its definition (9):

$$\int Q_{\alpha\beta}(r_1, r_2)\, dr_1\, dr_2 = N$$

$$\int Q_{\alpha\beta}(r_1, r_2)\, dr_1 = \rho_\beta(r_2).$$

The Hamiltonian is symmetric relative to the permutation of the replica indices; thus the one-replica values such as density $\rho_\alpha(r)$, do not depend on the replica index α. It seems natural to assume that the one-replica density $\rho_\alpha(r)$ will be the same as for a corresponding homopolymer with a two-particle interaction constant $\tilde{B}$ and a three-particle interaction constant C. It has been already mentioned that we shall treat the case when the chain is a globule, i.e. $\tilde{B} < 0$. The density of a large homopolymeric globule (with volume $V = N/\rho \gg a$) was found by Lifshitz *et al* (1978): $\rho_\alpha = -3\tilde{B}/2C$. The surface layer thickness is $a \ll V^{1/d}$ for $N \gg 1$. This now makes it possible to neglect the surface effect, i.e. to use the volume approximation. This approximation means that the average density for any replica $\rho_\alpha(r)$ is constant inside the globule and zero outside it. In such a globule the order parameter $Q_{\alpha\beta}(r_1, r_2)$ will depend only upon the difference between the arguments: $Q_{\alpha\beta}(r_1, r_2) = Q_{\alpha\beta}(r_1 - r_2)$ and the normalisation condition will have the form

$$\int Q_{\alpha\beta}(r)\, dr = \rho \tag{13}$$

($Q_{\alpha\beta}(r) = 0$ outside the globule).

We see that the averaged interaction constants B_0 and C are introduced into their theory through a single parameter—the density of a chain.

3. The mean-field approximation

We shall deal with (11) using the mean-field approximation, i.e. find the minimum of the integrand in (11) (in our case, it really demands maximisation instead of minimisation since when $n < 1$ both terms in the integrand of (11) change their sign).

The main problem now is the maximisation of the two-replica part of the free energy

$$F\{Q_{\alpha\beta}\} = -\tfrac{1}{2}B^2 V \int \sum_{\alpha \neq \beta} Q_{\alpha\beta}^2(r)\, dr - S\{Q_{\alpha\beta}(r)\}$$

together with the normalisation condition (13).

It is necessary to evaluate the configurational entropy $S\{Q_{\alpha\beta}\}$. The direct way to do this is to use the Lifshitz approach to the theory of globules (Lifshitz *et al* 1979). This is done in appendix 1 where it is shown that

$$\max_{Q_{\alpha\beta}} \{F\{Q_{\alpha\beta}\}\} = \max_{\Phi} \{F\{\Phi\}\}. \tag{14}$$

Here $\Phi(r_1, \ldots, r_n) = \Phi(\{r\})$ is the function of all replica coordinates

$$F\{\Phi\} = \left\{ \tfrac{1}{2}a^2 \int \sum_\alpha \left(\frac{d\Phi}{dr_\alpha}\right)^2 d\{r\} - \tfrac{1}{2}B^2 \int \sum_{\alpha \neq \beta} \Phi^2(\{r\})\Phi^2(\{r'\})\delta(r_\alpha - r'_\alpha) \right.$$

$$\left. \times \delta(r_\beta - r'_\beta)\, d\{r\}\, d\{r'\} \right\}. \tag{15}$$

Maximisation on the RHS of (14) must be done with the normalisation condition

$$\int \Phi^2(\{r\}) \, \mathrm{d}\{r\} = N. \tag{16}$$

$Q_{\alpha\beta}$ and Φ which maximise (14) are connected via the relation

$$Q_{\alpha\beta}(r_1, r_2) = \int \Phi^2(\{r'\})\delta(r_1 - r'_\alpha)\delta(r_2 - r'_\beta) \, \mathrm{d}\{r\}. \tag{17}$$

Variation of (15) gives

$$(a^2\Delta + \mu)\Phi = -2B^2\Phi(\{r\}) \int \sum_{\alpha \neq \beta} \Phi^2(\{r'\})\delta(r_\alpha - r'_\alpha)\delta(r_\beta - r'_\beta) \, \mathrm{d}\{r'\}. \tag{18}$$

Here Δ is a dn-dimensional Laplacian and μ is a Lagrangian multiplier which corresponds to the normalisation condition (16). We underline an important feature of function Φ which follows from translational invariance inside the large globule:

$$\Phi(r_1 + R, \ldots, r_n + R) = \Phi(r_1, \ldots, r_n) \tag{19}$$

for $R \ll V^{1/d}$.

4. Qualitative investigation

The exact maximisation of (15) for arbitrary n (especially for $n \to 0$) is a cumbersome problem. However, here we shall give simple arguments which make it possible to analyse the situation qualitatively. These arguments are analogous to the virial theorem in field theory (Rajaraman 1982).

We introduce a one-parameter set of Φ functions:

$$\Phi_R(\{r\}) = R^{d(1-n)/2}\Phi_1(\{r/R\}). \tag{20}$$

The scale R should itself be chosen from the maximisation condition. Φ_1 is the function with the unit scale. It can be seen easily that $F\{\Phi_R\}$ can be represented in a simple form:

$$F\{\Phi_R\} = A_1/R^2 - A_2/R^d \tag{21}$$

where

$$A_1 = \tfrac{1}{2}a^2 \int \sum_\alpha (\mathrm{d}\Phi_1/\mathrm{d}r_\alpha)^2 \mathrm{d}\{r\} \tag{22}$$

$$A_2 = \tfrac{1}{2}B^2 \int \sum_{\alpha \neq \beta} \Phi_1^2(\{r\})\Phi_1^2(\{r'\})\delta(r_\alpha - r'_\alpha)\delta(r_\beta - r'_\beta) \, \mathrm{d}\{r\} \, \mathrm{d}\{r'\}. \tag{23}$$

The values A_1 and A_2 do not depend on R and are both negative for $n < 1$. The problem of maximisation is split now into two problems:

(1) to find the scale R, i.e. to maximise (21) for $n < 1$ with respect to R;

(2) to find the form of Φ_1, i.e. to maximise (21) with respect to Φ_1.

The first can be done simply when $d < 2$. In this case we obtain $R = R^* = (2A_1/dA_2)^{1/(2-d)}$, so that $R \sim B^{-2/(2-d)}$. Thus the non-trivial solutions appear for any $B \neq 0$ in this case. The reverse scale $1/R^*$ is a continuous function of B ($1/R^* = 0$ for $B = 0$). We shall give a detailed investigation of the case $d < 2$ (the results for the case $d > 2$ will be published elsewhere).

Introducing the dimensionless variables

$$\tilde{r} = (\mu^{1/2}/a)r \qquad \text{and} \qquad \tilde{\Phi} = (\mu/a^2)^{(2-n)d/4}\Phi \qquad (24)$$

it is possible to exclude all parameters in (18). Taking the limit $n \to 0$ and applying the Parisi ansatz (Parisi 1980) to $Q_{\alpha\beta}(r)$ we shall obtain the function of two continuous arguments for the order parameter $Q(x, r)$ where $0 < x < 1$. This function obeys the following scaling relation:

$$Q(x; r) = s_1 g_d(x; r/s_2)$$

where $s_1 = \rho s_2^d$, $s_2 = (\rho B^2/a^2)^{1/(2-d)}$ and g_d is the universal scaling function. This means that the functional form of the order parameter, as well as the function Φ_1, is not changed when the parameters of the model B, B_0 and C are changed when the temperature or solvent quality is changed.

Φ_1 cannot be found by any perturbative method since it contains no small parameter after the rescaling. However, one can see that there is no solution for the minimisation equations (21)–(23) that gives symmetric $\Phi_1(r_1, \ldots, r_n)$. Let us consider, indeed, the large globule with $V \gg a^3$. It was mentioned above that in this case the function $\Phi_1(r_1, \ldots, r_n)$ is translationally invariant (see (19)). Therefore it can be represented in the form $\Phi_1(r_1 - r_n, r_2 - r_n, \ldots)$. It can be seen that the entropic term A_1 and energetic term A_2 in (21) are of different orders on n (the latter is usually $\sim n(n-1)$, while the former one is also negative but ~ 1). This fact signals that the existing solution for Φ_1 is necessarily non-symmetric which causes the symmetry breaking in the order parameter $Q_{\alpha\beta}(r)$ as well.

5. Gaussian approximation

An exact maximisation of (15) as well as a perturbative approach to it is impossible. It seems reasonable to use a variational approach with the simple Gaussian trial functions:

$$\Phi(\{r\}) = \Phi_0 \exp\left(-\tfrac{1}{2} \sum_{\alpha \neq \beta} k_{\alpha\beta} r_\alpha r_\beta\right) \qquad (25)$$

where $k_{\alpha\beta}$ is the matrix of the Parisi type.

The translational invariance of the function Φ leads to the condition for the matrix elements $k_{\alpha\beta}$:

$$\sum_\alpha k_{\alpha\beta} = 0. \qquad (26)$$

This condition together with the normalisation condition (16) gives

$$\Phi_0^2 = \pi^{(1-d)/2} \rho D^{d/2}. \qquad (27)$$

Here D is the minor of an arbitrary diagonal element of the matrix $k_{\alpha\beta}$; it does not depend on the choice of the element. Calculating the integrals in (15) with the use of (25)–(27) we obtain the final expression:

$$F\{\Phi\} = N\left(\frac{d}{4} a^2 \sum_\alpha k_{\alpha\alpha} - \frac{\rho B^2}{2(2\pi)^{d/2}} \sum_{\alpha \neq \beta} D_{\alpha\beta}^{-d/2}\right). \qquad (28)$$

Here $D_{\alpha\beta}$ is the determinant of the matrix obtained from $k_{\alpha\beta}$ by subtracting columns and rows which intersect in the diagonal elements $k_{\alpha\alpha}$ and $k_{\beta\beta}$, divided by D. We shall now use the Parisi ansatz and make the limit $n \to 0$. As a result we obtain

$$F\{\Phi\} = Nn \left(\frac{d}{4} a^2 \int_0^1 k(x) \, dx + \frac{\rho B^2}{2(2\pi)^{d/2}} \int_0^1 D^{-d/2}(x) \, dx \right) \qquad (29)$$

where $k(x)$ and $D(x)$ are the functions which are obtained as $n \to 0$ limit of $k_{\alpha\beta}$ and $D_{\alpha\beta}$. The function $D(x)$ is calculated in appendix 2 where it is shown that

$$D(x) = 2 \left(\frac{1}{xK(x)} - \int_x^1 \frac{dy}{y^2 K(y)} \right) \qquad (30)$$

where

$$K(x) = \int_0^x k(y) \, dy - xk(x). \qquad (31)$$

It is convenient now to express the functional (29) through the function $K(x)$:

$$F\{\Phi\} = \tfrac{1}{4} Nn \, da^2 \left(-\int_0^1 K(x) \, dx/x^2 + b^2 \int_0^1 D^{-d/2}(x) \, dx \right) \qquad (32)$$

where $b^2 = 2\rho B^2/(da^2(2\pi)^{d/2})$. Variation of (32) with respect to $K(x)$ gives

$$K^2(x) + db^2 \left(\int_0^x D^{-(d+2)/2}(y) \, dy - xD^{-(d+2)/2}(x) \right) = 0. \qquad (33)$$

Differentiation of (33) with respect to x gives

$$D(x) = [d(d+2)b^2/2]^{2/(d+4)} K^{6/(d+4)}(x) \qquad \text{or} \qquad K'(x) = 0. \qquad (34)$$

Differentiating (34) with respect to x we obtain finally

$$K(x) = K_0 x^m \qquad \text{or} \qquad K(x) = \text{constant}$$

where

$$m = (4+d)/(2-d) \qquad K_0 = [d(d+2)b^2/2]^{2/(2-d)}[3/(4+d)]^m.$$

The two solutions for $K(x)$ must coincide for $x = x_0$; (30) and (31) must be taken into account:

$$K(x) = \begin{cases} K_0 x^m & 0 < x < x_0 \\ K_0 x_0^m & x_0 < x < 1 \end{cases}$$

where $x_0 = (4+d)/6$.

The order parameter can be determined with the help of (17):

$$Q\{x, r\} = (\rho\pi^{-d/2}/R^d(x)) \exp[-(r/R(x))^2]$$

where

$$R^{-1}(x) = \begin{cases} (K_0/2x_0)^{1/2} x^{3/(2-d)} & 0 < x < x_0 \\ (K_0/2)^{1/2} x_0^{m/2} & x_0 < x < 1. \end{cases}$$

Finally, we obtain for the overlaps (10)

$$q(x) = (\pi^{-1/2} R^{-1}(x))^d.$$
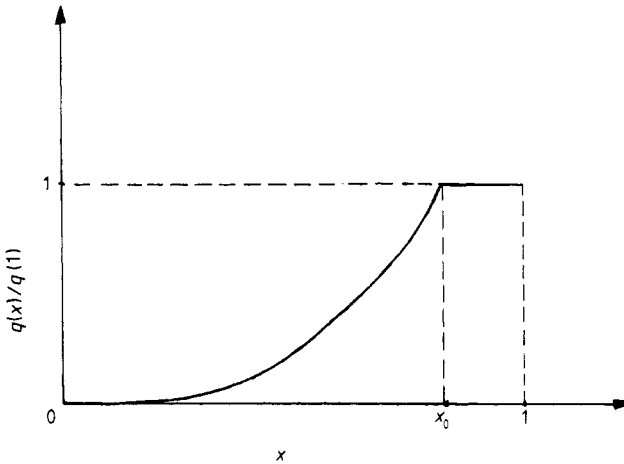
The plot of the function $q(x)$ is shown in figure 1.

Figure 1. Plot of the order parameter $q(x)$.

6. Discussion

The replica-symmetry breaking (RSB) means, as usual, the lack of ergodicity, i.e. the presence of a variety of equilibrium states (Parisi 1983). However, in the discussed system of a heteropolymer at a low space dimension $d < 2$ this RSB is quite unusual. The presence of a finite scale $R^* = R(1)$ of the order parameter $Q(x, r)$ means that the chain fold is frozen but not absolutely—only up to the scale R^*, i.e. the characteristic fluctuation of any link is of order R^*.

The number of equilibrium states (Mézard *et al* 1984), i.e. the shape of functions $Q(x, r)$ and $q(x)$ does not depend on the heterogeneity of the chain B, and only the equilibrium scale of fluctuations R^*, as well as the amplitude of the order parameter, depends on it. This means that the increase in heterogeneity leads to an increase in the degree of freezing in equilibrium states but without changing the nature of these states.

Another interesting and unusual feature of the RSB in this model is the power dependence $q \sim x^{3d/(2-d)}$ for $x < x_0$. Accepting the standard interpretation for the RSB (Parisi 1983) and introducing the function $P(q) = dx/dq$—the probability distribution for two states to have the overlap q—we obtain

$$P(q) \sim q^{(2-4d)/(3d)} \qquad \text{for} \qquad q < q(1).$$

The plot of function $P(q)$ is shown in figure 2. The rapid growth of $P(q)$ for $q \to 0$ shows an anomalously large number of weakly overlapping states.

It has already been mentioned that the short-range interactions are predominant for low-dimensional polymers. Thus, the polymeric bonds create anomalously strong frustrations which cause an anomalous growth of the number of states in heteropolymers at low $(d < 2)$ dimensions.

The power dependence of the order parameter $q(x)$ make it reasonable to suggest peculiar scaling properties of the set of equilibrium states. It seems possible to use the RG transformation of a simultaneous change of the scale of the order parameter and the scale of the replica variables x which means dividing replicas with similar (on the current scale) chain fold into clusters. Such an RG approach may reveal interesting fractal properties of the set of states.
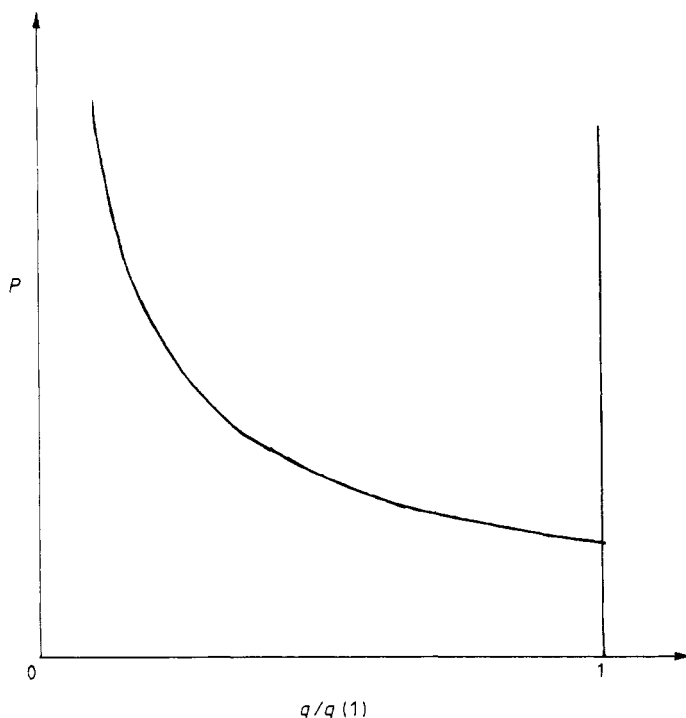
Figure 2. Plot of the probability distribution $P(q)$.

The question arises: what have the results for low dimensions to do with real proteins? It was stressed recently (Grosberg *et al* 1988a, b) that the topological constraints in a non-phantom polymer chain strongly influence the chain collapse process. This leads to the formation of a long-lived intermediate globular state—the fractal crumpled globule for which the short-ranged interactions are predominant even for $d = 3$. It may be possible that the suggested results may describe (at least, qualitatively) the structure of a crumpled globule.

Acknowledgments

Appendix 1. Evaluation of configurational entropy

We shall generalise the idea of the mean-field approach to the calculation of the configurational entropy proposed by Lifshitz *et al* (1978) for the case of homopolymeric globules. Consider the system of n replicas in an 'external field' $U_{\alpha\beta}(r_1, r_2)$. The partition function of such a system is

$$Z\{U_{\alpha\beta}\} = \exp(-F\{U_{\alpha\beta}\})$$

$$= \int Dr_\alpha(\tau) \exp\left(-\tfrac{1}{2}a^{-2}\int_0^N \sum_\alpha \left(\frac{dr_\alpha}{d\tau}\right)^2 d\tau - \int_0^N \sum_{\alpha\neq\beta} U_{\alpha\beta}(r_\alpha(\tau), r_\beta(\tau))\right).$$

The last integral can be expressed through integration with the order parameter $Q_{\alpha\beta}(r_1, r_2)$:

$$Z\{U_{\alpha\beta}\} = \int DQ_{\alpha\beta}(r_1, r_2)$$

$$\times \exp\left(-\int \sum_{\alpha \neq \beta} dr_1\, dr_2\, U_{\alpha\beta}(r_1, r_2)Q_{\alpha\beta}(r_1, r_2) + S\{Q_{\alpha\beta}\}\right). \qquad (A1.1)$$

The entropy $S\{Q_{\alpha\beta}\}$ in (A1.1) is the same as in (12). Since we are dealing with a globular state, it is reasonable to calculate the last integral by the saddle-point method. The saddle-point value of $Q_{\alpha\beta}$ is determined by the equation

$$U_{\alpha\beta}(r_1, r_2) = \delta S\{Q_{\alpha\beta}\}/\delta Q_{\alpha\beta}(r_1, r_2)$$

and the entropy

$$S\{Q_{\alpha\beta}\} = \int \sum_{\alpha \neq \beta} dr_1\, dr_2\, U_{\alpha\beta}(r_1, r_2)Q_{\alpha\beta}(r_1, r_2) - F\{U_{\alpha\beta}\}. \qquad (A1.2)$$

Thus the problem now is to calculate the free energy $F\{U_{\alpha\beta}\}$ of the n-replica system in the 'external field' $U_{\alpha\beta}(r_1, r_2)$. In order to do this we introduce the Green function which is the partition function of replicas with fixed ends:

$$G(0, \{x\}; N, \{y\}) = \int_{r_\alpha(0)=x_\alpha}^{r_\alpha(N)=y_\alpha} Dr_\alpha(\tau) \exp\left(-\tfrac{1}{2}a^{-2}\int_0^N \sum_\alpha \left(\frac{dr_\alpha}{d\tau}\right)^2 d\tau\right.$$

$$\left.-\int_0^N \sum_{\alpha \neq \beta} U_{\alpha\beta}(r_\alpha(\tau), r_\beta(\tau))\right) d\tau.$$

This Green function obeys the 'equation of motion':

$$\frac{dG}{dN} = \tfrac{1}{2}a^2 \sum_\alpha \frac{d^2 G}{dy_\alpha^2} - \sum_{\alpha \neq \beta} U_{\alpha\beta}(y_\alpha, y_\beta)G \qquad (A1.3)$$

for $U_{\alpha\beta} \ll 1$. The solution of (A1.3) has the following form in the ground-state dominance approximation:

$$G(0, \{x\}; N, \{y\}) = \exp(-\mu N)\Phi(\{x\})\Phi(\{y\}).$$

Here μ and Φ are the main eigenvalue and eigenfunction of the equation

$$\left(\tfrac{1}{2}a^2 \sum_\alpha \frac{d^2 G}{dy_\alpha^2 + \mu}\right)\Phi = \Phi \sum_{\alpha \neq \beta} U_{\alpha\beta}(y_\alpha, y_\beta). \qquad (A1.4)$$

The free energy has the simple form

$$F\{U_{\alpha\beta}\} = N\mu \qquad (A1.5)$$

and overlaps

$$Q_{\alpha\beta}(r_1, r_2) = \delta F\{U_{\alpha\beta}\}/\delta U_{\alpha\beta}(r_1, r_2)$$

$$= \int \Phi^2(\{y\})\delta(y_\alpha - r_1)\delta(y_\beta - r_2)\, d\{y\}. \qquad (A1.6)$$

Substituting (A1.4)-(A1.6) into (A1.2) we obtain

$$S\{Q_{\alpha\beta}\} = \tfrac{1}{2}a^2 \int \sum_\alpha \left(\frac{d\Phi}{dy_\alpha}\right)^2 d\{y\}. \qquad (A1.7)$$

Equations (15) and (17) follow from (A1.6) and (A1.7).

Appendix 2. Evaluation of the function $D(x)$

Consider matrix M obtained from $k_{\alpha\beta}$ by crossing out the row and the column of $k_{\alpha\beta}$ which intersect in the αth diagonal element of $k_{\alpha\beta}$. It can be seen that the βth diagonal element of the reverse matrix M^{-1} is $D_{\alpha\beta}$. Thus it is necessary to evaluate the matrix M^{-1} in the limit $n \to 0$.

In the framework of the Parizi ansatz, the matrix M is characterised by the sequences m_i and k_i—the size of blocks and values of the matrix elements inside the blocks (figure $3(a)$). This matrix is expressed through the function of a continuous argument $k(x)$, $0 < x < 1$, in the limit $n \to 0$. The reverse matrix M^{-1} has a lower symmetry and it is necessary to determine, besides, m_i the values D_i, E_{ij} and F_{ij}; the meaning of these values is clear from figure $3(b)$.

The relation $M^{-1}M = 1$ leads to the system of linear equations for D_i, E_{ij} and F_{ij}. Taking the limit $n \to 0$ we obtain the following integral equations:

$$\int_0^y k(z)F(y, z)\,dz + \int_y^x k(z)F(z, y)\,dz + k(x)\int_x^1 F(z, y)\,dz + k(y)\Big(yE(y, y)$$

$$+ \int_y^1 E(z, y)\,dz - D(y)\Big) - K(x)F(x, y) = 0$$

$$\int_0^y k(z)F(y, z)\,dz + k(y)\int_y^1 F(z, y)\,dz - k(x)D(y) + yk(y)E(y, y) + \int_y^x k(z)E(z, y)\,dz$$

$$+ k(x)\int_x^1 E(z, y)\,dz - K(x)E(x, y) = 0$$

$$\int_0^y k(z)F(y, z)\,dz + k(y)\int_y^1 F(z, y)\,dz - D(y)\int_0^1 k(z)\,dz + yk(y)E(y, y)$$

$$+ \int_y^1 k(z)E(z, y)\,dz = -1.$$

Here $D(y)$, $E(x, y)$ and $F(x, y)$ are the functions of continuous arguments $0 < y < 1$, $y < x < 1$; $K(x)$ is determined by equation (31) in the text.

The solution of this system is obtained by differentiation of the equations. As a result we obtain
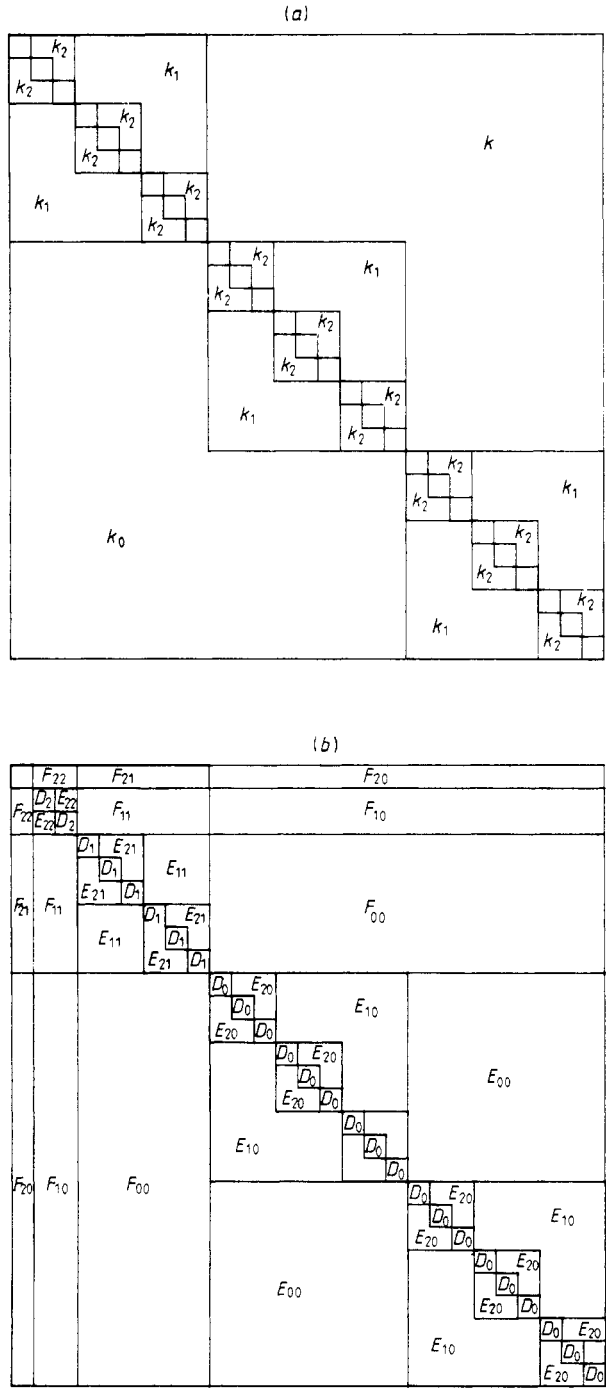
$$F(x, y) = \frac{1}{xK(x)} - \int_x^1 \frac{dz}{z^2 K(z)} \tag{A2.1}$$

$$E(x, y) = D(y) - F(x, y) \tag{A2.2}$$

$$D(y) = 2F(y, y). \tag{A2.3}$$

Equation (30) of the text is obtained from (A2.1) and (A2.3).

Note added. This paper had been already submitted when we read an interesting work (Garel and Orland 1988) in which similar ideas have been put forth. Although close in general approach, our work differs from that of Garel and Orland (GO) in the character of the results obtained. For example, it has been argued in GO that the discontinuous Potts-glass-like behaviour of the order parameter Q is obtained for high dimensions, $d \gg 4$. This is in contrast with our result in §4 of this paper where we claim that the discontinuous behaviour is observed when $d > 2$ (see also Shakhnovich and Gutin 1989). For low dimensions the use of the multiplicative replica-symmetric trial function φ in GO made it possible to consider the coil-disordered globule transition, while the main aim of the present work was the investigation of the non-ergodic frozen globular state with broken replica symmetry. This caused the choice of the trial Ψ function $\check{\Psi}(\{R\}) = \Psi_0 \exp(-\frac{1}{2}\Sigma_{\alpha,\beta} k_{\alpha\beta}R_\alpha R_\beta)$ in which replica-symmetry breaking is assumed.

Figure 3. (*a*) Matrix *k*. (*b*) Matrix M^{-1}. (Details in appendix 2.)

References

Anfinsen C B 1973 *Science* **181** 223
Binder K and Young A P 1986 *Rev. Mod. Phys.* **58** 801
Bryngelson J B and Wolynes P G 1987 *Proc. Natl Acad. Sci. USA* **84** 7524
Creighton T E 1985 *J. Phys. Chem.* **89** 2452
Derrida B 1980 *Phys. Rev. Lett.* **45** 79
Edwards S F 1966 *Proc. R. Soc.* **88** 266
Fernandez A and Rabitz H 1987 *Biophys. Chem.* **28** 89
Garel T and Orland H 1988 *Europhys. Lett.* **6** 307
Grosberg A Yu, Nechaev S K and Shakhnovich E I 1988a *Biofisika* **33** 247
—— 1988b *J. Physique* **49** 2095
Kholodenko S K and Freed K F 1983 *J. Chem. Phys.* **78** 7390
Lifshitz I M, Grosberg A Yu and Khokhlov A R 1978 *Rev. Mod. Phys.* **50** 683
Mézard M G, Parisi G, Sourlas N, Toulouse G and Virasoro M 1984 *J. Physique* **45** 843
Obukhov S P 1986 *J. Phys. A: Math. Gen.* **19** 3655
Paine G H and Scheraga H A 1985 *Biopolymers* **24** 1391
Parisi G 1980 *J. Phys. A: Math. Gen.* **13** L155, 1101, 1887
—— 1983 *Phys. Rev. Lett.* **50** 1946
Rajaraman R 1982 *Solitons and Instantons* (Amsterdam: North-Holland)
Rammal R, Toulouse G and Virasoro M 1986 *Rev. Mod. Phys.* **52** 765
Shakhnovich E I and Gutin A M 1989 *Europhys. Lett.* in press
Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **32** 1792