Globular state of branched random heteropolymers

1993 J. Phys. A: Math. Gen. 26 1037

(http://iopscience.iop.org/0305-4470/26/5/023)

J. Phys. A: Math. Gen. 26 (1993) 1037–1049. Printed in the UK

Globular state of branched random heteropolymers

A M Gutin, A Yu Grosberg† and E I Shakhnovich‡
Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

Received 8 May 1992, in final form 16 October 1992

Abstract. The globular state of a randomly branched disordered heteropolymer—the simplest model of RNA with given secondary structure—is investigated using the replica approach in the mean-field approximation. It is shown that for dimensions of space less than four there is replica-symmetry breaking with power behaviour of the Parisi order parameter. The scale of fluctuations of links decreases when heterogeneity of interaction in the molecule increases.

1. Introduction

Statistical mechanics of heteropolymers is very interesting due to obvious biophysical applications for investigation of proteins and nucleic acids. Several statistical-mechanical models of proteins were suggested recently and the results are encouraging (Bryngelson and Wolynes 1987, Garel and Orland 1988, Shakhnovich and Gutin 1989a, c, 1990a; see the review in Karplus and Shakhnovich 1992). In most of these papers a simple model of the protein chain—linear heteropolymer with disordered sequence of monomers—has been investigated. Powerful techniques developed in the spin-glass theory (Binder and Young 1986) were used in these studies and considerable progress was achieved. It was shown (Shakhnovich and Gutin 1989b) that statistical mechanics of random linear heteropolymer is very sensitive to space dimensionality d making heteropolymers at $d > 2$ dramatically different from the low-dimensional ($d \leqslant 2$) case. In the latter case there is unusual replica symmetry breaking with power dependence of the Parisi order parameter and ultrametric structure of configurational space.

In the present paper we study equilibrium properties of another biological macromolecule—RNA. The principal difference of RNA molecules from proteins is that, after formation of secondary structure with several short helices, RNA molecules possess a tree-like structure (see for example, the model of secondary structure of 16S rRNA (Gutell *et al* 1985). Though the positions of branch points are determined in principle by minimization of all interactions involved (as in proteins secondary structure is strongly influenced by tertiary structure) this may be not relevant for RNA where a complementarity rule dictates strong pairing and corresponding hydrogen bonds make the secondary structure pattern stable and fixed even in the absence of tertiary interactions. This point of view is favoured by experimental

† Present address: Institute of Chemical Physics, Russian Academy of Sciences, Moscow, Russia.
‡ Author to whom any correspondence should be addressed.

0305-4470/93/051037+13$07.50

results on tRNA melting which demonstrated the high stability of secondary structure (clover leaf) after unfolding of the tertiary structure (Privalov and Filimonov 1978). Calculation of energies of secondary structures in random and natural RNA (Higgs 1992) demonstrates that these energies are significant (tens of kcal m^{-1}) and can really 'quench' secondary structure. This makes it reasonable to consider *quenched branching* as a starting approximation effect of secondary structure. Therefore our model of RNA is not linear but a randomly branched molecule with helices as monomers. Intramolecular interactions in a chain depend on the chemical structure of interacting monomers: therefore we have to consider the randomly branched disordered heteropolymer. As the formation of secondary structure chains stiffens, long-range interactions occur not between monomers but rather between parts of the chain. This means that though initially there were only four types of monomers in RNA this 'renormalization' of monomers effectively increases the number of their types, which makes energies of interactions statistically independent. These models for linear heteropolymers were introduced in Garel and Orland (1989) and Shakhnovich and Gutin (1989a–c).

The model is twice disordered because we have (a) a random quenched structure of branching and (b) random energies of interactions between monomers. The latter disorder has been studied for linear heteropolymers in Shakhnovich and Gutin (1989a–c). As for randomness caused by branching, it has been much less frequently investigated it is much less investigated and there is even no clear distinction between annealed and quenched cases (see, for example, Daoud *et al* (1983), Gaunt and Flesia (1991), Gutin *et al* (1992)). It should be noted that the methods used in the present paper are similar to the those suggested in Shakhnovich and Gutin (1989b) and applied for other systems in Mezard and Parisi (1992a, b). However, the crucial difference with the case of a linear polymer is in configurational entropy caused by polymeric bonds. Branched polymers are more compact and many contacts between monomers are formed by chain neighbours so that the polymeric structure of a branched molecule plays a very significant role in its thermodynamic behaviour. This difference leads to power dependence of the Parisi order parameter for $d < 4$ and hence, to a completely different structure of configurational space. Therefore, in real three-dimensional space there is a dramatic difference in the equilibrium behaviour of molecules of RNA and proteins exhibit dramatic difference because proteins (linear heteropolymers) can be described (Shakhnovich and Gutin 1989c, 1990) by the random energy model (REM) (Derrida 1980) with different low-energy states being essentially different in structure and therefore not correlated. As we shall show in this paper, this is not the case for branched polymers and the strong influence of polymeric bonds in this case causes significant effects that makes configurational space ultrametric.

2. The model and basic relations

The partition function of a branched molecule can be presented in the form

$$Z_{B,v} = \int \prod_{i=1}^{N} \mathrm{d}x_i \, G_B\{x_i\} \exp\left(-\frac{1}{2T} \sum_{i \neq j} v_{ij} \delta(x_i - x_j) - \frac{1}{6T} W \sum_{i \neq j \neq k} \delta(x_i - x_j)\delta(x_i - x_k) \right) \tag{1}$$

where integration is taken over space coordinates x_i of all monomers ($i = 1, 2, \ldots, N$), T is temperature (we use units in which the Boltzmann constant $k_B = 1$), v_{ij} is the two-body interaction constant between ith and jth monomers and W is a three-body interaction constant which is assumed to be independent of the type of monomers. The factor G_B depends on the coordinates of all monomers $\{x_i\}$ and it describes the structure of branchings of the molecule

$$G_B\{x_i\} = \prod_{(i,j)} g(x_i - x_j) \tag{2}$$

where the product is taken only over pairs of monomers (i, j) that are chemically bonded and the factor g in its simplest form is Gaussian (Lifshitz *et al* 1978):

$$g(x) = (2\pi a^2/d)^{-d/2} \exp\left(-dx^2/2a^2\right). \tag{3}$$

Here a is the mean distance between the neighbouring monomers. The two-body interaction constants v_{ij} are assumed to be independent random variables with Gaussian distribution

$$P(v_{ij}) = (2\pi v^2)^{-d/2} \exp\left(-(v_{ij} - \bar{v})^2/2v^2\right) \tag{4}$$

where d is the dimension of space, $\bar{v}$ and v are the mean and the standard variance respectively. This condition of statistical independence and Gaussian distribution of interaction parameters v_{ij} may seem too restricting for RNA as the number of types of monomers there is only four. However, the recent solution for the freezing transition in a 'two-letter' heteropolymer, where interaction energies are by no means independent, demonstrated similarity in the thermodynamic behaviour of such heteropolymers with models with independent interactions (Sfatos *et al* 1992). This justifies taking interaction energies in the simplest form (4).

We have to evaluate the average free energy using the replica trick

$$\langle\langle \ln Z_{v,B}\rangle_v\rangle_B \geqslant \lim_{n\to 0} \frac{\langle\langle Z^n_{v,B}\rangle_v\rangle_B > -1}{n} \tag{5}$$

where $\langle\cdots\rangle_v$ denotes averaging over disorder caused by heterogeneity of monomers and $\langle\cdots\rangle_B$ denotes averaging over random structures of branching. Averaging over v_{ij} with weight (3) gives

$$\langle\langle Z^n_{v,B}\rangle_v\rangle_B \geqslant \int \prod_{i=1}^{N}\prod_{\alpha=1}^{n} \mathrm{d}x_i^\alpha\, G_n\{x_i^\alpha\} \exp\left(-\frac{1}{T} H_{\mathrm{eff}}\{x_i^\alpha\}\right) \tag{6}$$

where integration is taken over coordinates of all monomers of all n replicas $\{x_i^\alpha\}$ and

$$H_{\mathrm{eff}}\{x_i^\alpha\} = \tfrac{1}{2} v_{\mathrm{eff}} \sum_\alpha \sum_{i\neq j} \delta(x_i^\alpha - x_j^\alpha) + \tfrac{1}{6} W \sum_\alpha \sum_{i\neq j\neq k} \delta(x_i^\alpha - x_j^\alpha)\delta(x_i^\alpha - x_k^\alpha)$$
$$- \tfrac{1}{2} v^2 \sum_{\alpha\neq\beta} \sum_{i\neq j} \delta(x_i^\alpha - x_j^\alpha)\delta(x_i^\beta - x_j^\beta) \tag{7}$$

where $v_{\text{eff}} = \bar{v} - v^2/2u$ where u is a monomer volume which corresponds to the scale of δ-functions in the definition of the Hamiltonian of the system. This growth of average attraction due to disorder has been discussed in (Shakhnovich and Gutin 1989a, Higgs and Joanny 1991, Stepanow *et al* 1992). This effect comes from terms with $\alpha = \beta$ and exists even in the annealed case. It corresponds physically to the fact that heterogeneity leads to some rearrangement of a molecule (even without freezing) when monomers which attract each other become closer in space. This rearrangement is facilitated in more compact molecules and effective attraction occurs.

$$G_n\{x_i^\alpha\} = \left\langle \prod_{\alpha=1}^{n} G_B\{x_i^\alpha\} \right\rangle_B . \tag{8}$$

We now define the macroscopic order parameter $\rho(X)$ which depends on the dn-dimensional vector $X = (X_1, \ldots, X_n)$

$$\rho(X) = \sum_{i=1}^{N} \prod_{\alpha=1}^{n} \delta(x_i^\alpha - X_\alpha). \tag{9}$$

Expression (6) can be rewritten in terms of ρ

$$\langle\langle Z_{v,B}^n \rangle_v \rangle_B \geqslant \int D\rho(X) \exp\left(-\tfrac{1}{T} E(\rho) + S(\rho)\right) \tag{10}$$

where energy term E is given by

$$E(\rho) = \frac{1}{2} v_{\text{eff}} \sum_\alpha \int n_\alpha^2(r)\mathrm{d}r + \frac{1}{6} W \sum_\alpha \int n_\alpha^3(r)\mathrm{d}r - \frac{1}{2} v^2 \sum_{\alpha \neq \beta} \int Q_{\alpha\beta}^2(r_1, r_2)\mathrm{d}r_1 \mathrm{d}r_2 \tag{11}$$

where

$$n_\alpha(r) = \sum_i \delta(x_i^\alpha - r) = \int \rho(X)\delta(X_\alpha - r)\mathrm{d}X \tag{12}$$

is density of monomers of replica α and

$$Q_{\alpha\beta}(r_1, r_2) = \sum_i \delta(x_i^\alpha - r_1)\delta(x_i^\beta - r_2) = \int \rho(X)\delta(x_\alpha - r_1)\delta(x_\beta - r_2) \tag{13}$$

is the correlator of replicas α and β. We can also introduce a more physically transparent parameter which has the meaning of structural overlaps, or degree of similarity between replicas

$$q_{\alpha\beta} = \frac{1}{N} \sum_i \delta(x_i^\alpha - x_i^\beta) = \frac{1}{N} \int Q_{\alpha\beta}(r, r)\mathrm{d}r. \tag{14}$$

The entropy term $S(\rho)$ is given by

$$S(\rho) = \ln \int \prod_{i=1}^{N} \prod_{\alpha=1}^{n} \mathrm{d}x_i^\alpha \delta \left[\rho(X) - \sum_{i=1}^{N} \prod_{\alpha=1}^{n} \delta(x_i^\alpha - x_\alpha) \right] G_n\{x_i^\alpha\} \tag{15}$$

and corresponds to the number of conformations of all replicas which have a given $\rho(X)$.

3. Entropy of the branched globule and the mean-field approximation

We are interested in the globular state of the molecule, and can use a mean-field approximation to determine the free energy and the equilibrium order parameter. In this approximation the replica free energy is

$$F\{\rho\} = E\{\rho\} - TS\{\rho\}. \tag{16}$$

In order to evaluate entropy S we use the method proposed by Lifshitz (Lifshitz *et al* 1978) for the case of a globule formed by a linear homopolymer. Consider the system of n replicas in an external field. The partition function of such a system is

$$Z_n\{\rho\} = \int \prod_{\alpha=1}^{n} \prod_{i=1}^{N} \mathrm{d}x_i^\alpha \, G_n\{x_i^\alpha\} \exp\left(-\frac{1}{T}\sum_{i=1}^{N} \phi(X_i)\right). \tag{17}$$

It can be written as an integral over order parameter ρ

$$Z_n\{\phi\} = \int D\rho(X) \exp\left(-\frac{1}{T}\int \phi(X)\rho(X)\,\mathrm{d}X + S\{\rho\}\right). \tag{18}$$

In the mean-field approximation this integral is calculated by the saddle-point method with the result

$$S\{\rho\} = \frac{1}{T}\int \phi(X)\rho(X)\,\mathrm{d}X + \ln Z_n\{\phi\} \tag{19}$$

where saddle-point of ρ is determined by the equation

$$\phi(X) = T\frac{\delta S}{\delta \rho(X)}. \tag{20}$$

Therefore the order parameter is

$$\rho(X) = -T\frac{\delta \ln Z_n\{\phi\}}{\delta \phi(X)}. \tag{21}$$

Now we have to evaluate the partititon function $Z_n\{\phi\}$ of n-replica system in the external field. In order to do this we introduce $Z_N(X)$ as follows:

$$Z_N(X) = M \int \prod_{\alpha=1}^{n} \prod_{i=1}^{N} \mathrm{d}x_i^\alpha \, G_n\{x_i^\alpha\} \exp\left(-\frac{1}{T}\sum_{i=1}^{N-1} \phi(X_i)\right)\delta(X_N - X) \tag{22}$$

where M is the number of all possible structures of branching. Taking into account the definition (8) it is possible to rewrite (22) in the form

$$Z_N(X) = \int \prod_{\alpha=1}^{n} \prod_{i=1}^{N-1} \mathrm{d}x_i^\alpha \left(\sum_B \prod_{\alpha=1}^{n} G_n\{x_i^\alpha\}\right) \exp\left(-\frac{1}{T}\sum_{i=1}^{N} \phi(X_i)\right)\delta(X_N - X) \tag{23}$$

where $\sum_B$ denotes summation over all possible structures of branchings ($M = \sum_B 1$).

We will consider now the simplest case (the generalization is straightforward) when each monomer can belong to one of two types: (a) the end monomer which is bonded only with one other monomer and (b) the branching monomer which is bonded exactly with three other monomers (figure 1). Moreover we suppose that the molecule is tree-like, i.e. without closed cycles. In this case the number of monomers N is even. A recursion formula for $Z_N(X)$ can be written (see, for details, deGennes (1968))

$$Z_N(X) = \int \mathrm{d}X' \prod_{\alpha=1}^{n} g(x_\alpha - x'_\alpha) \exp\left(-\frac{1}{T}\phi(X')\right) \times \sum_{2 \leqslant m \leqslant N-2} Z_m(X') Z_{N-m}(X'). \tag{24}$$

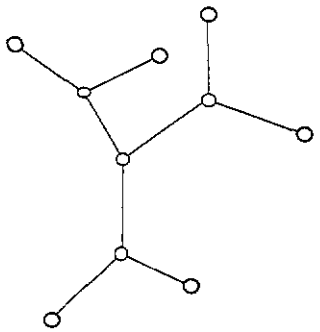

Figure 1. The model of a branched chain.

The meaning of this recursion equation is that the end monomer with coordinates X is bonded with branching which is bonded with two parts of the molecule having m and $(N-m)$ monomers. Moreover monomers at branching points play a role of end monomers for both remaining parts of the molecule. It should be noted that (24) is valid when $N \geqslant 4$. For $N = 2$ we have

$$Z_2 = \hat{g}(\exp(-\phi/T)) \tag{25}$$

where we introduce operator $\hat{g}$ as follows:

$$\hat{q}f = \int \prod_{\alpha=1}^{n} g(r_\alpha - r'_\alpha) f(r') \, \mathrm{d}r'. \tag{26}$$

The recursion relation (24) can be simplified by the introduction of a generating function $Z_p(X) = \sum_{N=2}^{\infty} Z_N(X) p^N$

$$Z_p = \hat{g}\left(\exp(-\phi/T)(p^2 + Z_p^2)\right). \tag{27}$$

It is well known that in the thermodynamic limit $N \to \infty$ all that is important for the partition function $Z_N(X)$ are the singularities of the generation function as a

function of the complex variable p. The equation (27) is quadratic. Therefore we assume that it's solution has square root singularity (see also DeGennes (1968))

$$Z_p(X) = \psi(X)(p^* - p)^{1/2} + \text{regular part} \tag{28}$$

and in thermodynamic the limit $N \to \infty$ we have

$$Z_n\{\phi\} = M^{-1} Z_N(X) \approx (p^*)^{-N} M^{-1}. \tag{29}$$

Differentiating (27) with respect to p and isolating the singularity when $p \to p^*$ as a result we have

$$\psi = 2\hat{g}\Big(\exp(-\phi/T) Z \psi\Big) \tag{30}$$

where $Z(X) = \lim_{p \to p^*} Z_p(X)$. Finally, when $p = p^*$ the equation (27) can be written in the following form:

$$Z = \hat{g}(\exp(-\phi/T)(p^{*2} + Z^2)). \tag{31}$$

Now we may determine the order parameter $\rho(X)$. In order to do this we use equations (21) and (29) with the result

$$\rho(X) = NT\delta \ln p^*/\delta\phi(X). \tag{32}$$

Variation of (31) with respect to $\phi(X)$ taking into account (30) gives

$$\rho = \tfrac{1}{2}\phi \exp(-\phi/T)(1 + Z^2/p^{*2}). \tag{33}$$

Expressing the field ϕ from this equation and substituting into (30), (31) and (19) we obtain finally

$$S\{\rho\} = \int \mathrm{d}X\, \rho(X) \ln\left(\tilde{Z}\frac{\hat{g}\Psi}{\Psi}\right) \tag{34}$$

where function $\Psi(X)$ is the solution of equation

$$\rho = \frac{\Psi}{2}(\tilde{Z} + \tilde{Z}^{-1}) \tag{35}$$

and

$$\tilde{Z} = \hat{g}(\rho/\hat{g}\Psi). \tag{36}$$

This very cumbersome expression for the entropy can be simplified when the n-replica density $\rho(X)$ is a 'smooth' function of space variables so that its characteristic scale R is larger than the mean length of a polymeric bond a ($R \gg a$). In this case the operator $\hat{g}$ up to the second order in (a/R) is

$$\hat{g} = 1 + a^2\Delta + \tfrac{1}{2}a^4\Delta^2 \tag{37}$$

where

$$\Delta = \sum_{\alpha=1}^{n} \partial^2/\partial x_\alpha^2. \tag{38}$$

Solving equation (35) and then calculating the entropy (34) to the same order, we obtain a simple result

$$S\{\rho\} = -\frac{1}{2}a^4 \int \frac{(\Delta\rho)^2}{\rho}\,\mathrm{d}X. \tag{39}$$

It should be noted that the obtained expression for entropy is substantially different from the entropy of a linear polymer (cf equation (A1.7) of Shakhnovich and Gutin (1989b)). Indeed the entropy of a branched molecule is of the order of $(a/R)^4$ while the entropy of the linear molecule is of the order of $(a/R)^2$. What is even more important, in this expression for entropy quenched character of branchings, is reflected explicitly. Indeed the nonlinearity in the Laplace operator (which, as given by (38) contains differentiation with respect to coordinates of all replicas) means that interaction between replicas (where α, β terms mix) show up in entropy. This is a direct consequence of quenched disorder in branchings (we remind the reader that in a linear polymer the effect of quenched disorder, as revealed by interreplica interactions, enters only on the level of averaging of interaction energy (Garel and Orland 1989, Shakhnovich and Gutin 1989a).

Nevertheless, as in the case of linear molecules, for branched molecules entropy $S\{\rho\}$ vanishes when the order parameter ρ is constant (more precisely when $\Delta\rho = 0$). This fact allows us to consider a globule formed by a large molecule in the volume approximation (Lifshitz *et al* 1978) neglecting surface effects. In the volume approximation the density of monomers of any replica n_α is constant inside the large globule, being independent of the replica index, and is determined by one-replica terms of energy (11)

$$n_\alpha = n_0 = -2v_{\mathrm{eff}}/3w. \tag{40}$$

Therefore in the volume approximation in order to determine a non-trivial order parameter $Q_{\alpha\beta}$ it is necessary to maximize replica free energy

$$F\{\rho\} = -\frac{1}{2}v^2 \sum_{\alpha\neq\beta} \int \mathrm{d}r_1\,\mathrm{d}r_2\, Q_{\alpha\beta}^2(r_1, r_2) - TS\{\rho\} \tag{41}$$

with normalization condition

$$\int \rho(X)\delta(X_\alpha - r)\,\mathrm{d}X = n_0(r). \tag{42}$$

It should be noted that order parameter $\rho(X)$ is translationally invariant inside the large globule, i.e.

$$\rho(x_1 + x, x_2 + x, \cdots, x_n + x) = \rho(x_1, x_2, \cdots, x_n) \tag{43}$$

for any $x^d \ll V$ where V is the volume of the globule.

4. Qualitative analysis

First of all we analyse qualitatively possible regimes of the system.

(1) Assume that the characteristic scale of the order parameter $\rho(X)$ is R, then as it follows from the normalization condition (42) the energy term in (41) $E \sim Nv^2 n_0 R^{-d}$ and the entropy term $S \sim N(a/R)^4$. Therefore for $d \leqslant 4$ free energy F as a function of R has it's maximum at

$$R \sim R^* = (Ta^4/n_0 v^2)^{1/(4-d)}. \tag{44}$$

Characteristic scale R^* increases when molecule heterogeneity v decreases. ($R^* \to \infty$ at $v \to 0$). We shall not consider the case $d > 4$.

(2) Then we introduce a dimensionless variable y via the relation: $X = R^* y$; it allows one to represent free energy in dimensionless and temperature-independent form

$$f = \frac{F}{v^2 n_0^2} = -\frac{1}{2} \sum_{\alpha \neq \beta} \int \mathrm{d}y \, \mathrm{d}y' \, \nu(y)\nu(y')\delta(y_\alpha - y'_\alpha)\delta(y_\beta - y'_\beta) + \frac{1}{2} \int \mathrm{d}y \frac{(\Delta\nu)^2}{\nu} \tag{45}$$

and normalization condition

$$\int \nu(y)\delta(y_\alpha - r)\mathrm{d}y = 1. \tag{46}$$

Now it is clear that the equilibrium form of the order parameter as well as the nature of replica symmetry breaking do not depend on parameters of the model $\bar{v}, v, w$ and a. All the dependence on these parameters is introduced via the dependence characteristic scale R^* on these parameters.

(3) The translational invariance of the order parameter $\rho(x)$ inside the globule causes the replica symmetry breaking. Indeed in the opposite case the energy term in (41) is of the order n (in the limit $n \to 0$) while the entropy is of order 1.

5. Gaussian variational procedure

An exact maximization of the free energy F given by the expression (41) is impossible. In this situation it seems reasonable to use a variational approach with the simple Gaussian trial functions Shakhnovich and Gutin (1989b).

$$\rho(X) = \rho_0 \exp\left(-\sum_{\alpha,\beta} k_{\alpha\beta} x_\alpha x_\beta \right) \tag{47}$$

where $k_{\alpha\beta}$ is the trial matrix which would reflect a possible replica-symmetry breaking; we will assume that it is of the Parisi type.

The translational invariance of the order parameter (43) leads to the condition for the matrix elements $k_{\alpha\beta}$

$$\sum_\alpha k_{\alpha\beta} = 0. \tag{48}$$

This condition together with the normalization condition (42) gives

$$\rho_0 = n_0(D\pi^{(1-n)})^{d/2} \tag{49}$$

where D is the minor of an arbitrary diagonal element of the matrix $k_{\alpha\beta}$.

The choice of the trial function in the form (47) has the advantage that integrals in (41) are Gaussian and can be evaluated using (48) and (49). The energy term in (41) is exactly the same as in Shakhnovich and Gutin (1989b) while entropy is different and it is entropic term which makes a difference between cases of branched and linear polymers. Taking ρ in the form (47) and substituting it to (39) we obtain

$$S = -4dNa^4 \sum_{\alpha\beta} k_{\alpha\beta}^2 = -4dNa^4 \left(\sum_{\alpha} k_{\alpha\alpha}^2 + \sum_{\alpha \neq \beta} k_{\alpha\beta}^2 \right). \tag{50}$$

Correspondingly, for free energy we have

$$F\{\rho\} = N\left(4da^4 \sum_{\alpha\beta} k_{\alpha\beta}^2 - \frac{n_0 v^2}{2(2\pi)^{d/2}} \sum_{\alpha \neq \beta} D_{\alpha\beta}^{-d/2} \right) \tag{51}$$

where $D_{\alpha\beta}$ 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 use now Parisi ansatz (Parisi 1980) and make the limit $n \to 0$. Instead of the matrices $k_{\alpha\beta}$ and $D_{\alpha\beta}$ we have now functions $k(x)$ and $D(x)$ respectively which are defined at $0 \leqslant x \leqslant 1$. In terms of these functions the free energy (51) has the form

$$F\{\rho\} = Nn\left[4da^4 \left(\left[\int_0^1 k(x)\,\mathrm{d}x \right]^2 - \int_0^1 k^2(x)\,\mathrm{d}x \right) + \frac{n_0 v^2}{2(2\pi)^{d/2}} \int_0^1 D(x)^{-d/2}\,\mathrm{d}x \right]. \tag{52}$$

The function $D(x)$ is calculated in appendix 2 of Shakhnovich and Gutin (1989b) with the result

$$D(x) = 2\left(\frac{1}{xK(x)} - \int_x^1 \frac{\mathrm{d}y}{y^2 K(y)} \right) \tag{53}$$

where

$$K(x) = \int_0^x k(y)\,\mathrm{d}y - xk(x). \tag{54}$$

Variation of (52) with respect to $K(x)$ gives

$$16(2\pi)^{d/2}(R^*)^{4-d}K^3(x) = xD^{-(d+2)/2}(x) - \int_0^x yD^{-(d+2)/2}(y)\,\mathrm{d}y. \tag{55}$$

Differentiating twice this equation with respect to x we obtain finally

$$K(x) = \begin{cases} K_0 x^m & \text{if } 0 \leqslant x \leqslant x_0 \\ K_0 x_0^m & \text{if } x_0 \leqslant x \leqslant 1 \end{cases} \tag{56}$$

where

$$x_0 = \tfrac{1}{8}(d+4) \qquad m = (4+d)/(4-d)$$
$$K_0 = \left(\frac{d+4}{4}\right)^{-m} \left(\frac{48(2\pi)^{d/2}}{d+2}\right)^{2/(d-4)} (R^*)^{-2}.$$

It is easy now to calculate the correlator of replicas $Q_{\alpha\beta}(r_1, r_2)$ defined by (13). Translational invariance of the order parameter implies that the correlator depends only upon the difference of the space arguments: $Q_{\alpha\beta}(r_1, r_2) = Q_{\alpha\beta}(r_1 - r_2)$. Taking the limit $n \to 0$ we obtain the function of two variables $Q(x; r)(0 \leqslant x \leqslant 1)$ which is determined by the expression

$$Q(x; r) = (n_0 \pi^{-d/2} R^{-d}(x)) \exp\left(-(r/R(x))^2\right) \tag{57}$$

where

$$R^{-1}(x) = \begin{cases} (K_0/2x_0)^{1/2} x^{4/(4-d)} & \text{if } 0 \leqslant x \leqslant x_0 \\ (K_0/2)^{1/2} x_0^{m/2} & \text{if } x_0 \leqslant x \leqslant 1. \end{cases} \tag{58}$$

It should be noted that $R(x_0)$ coincides with R* up to numerical factor.

We may express our result in terms of the physically more clear parameter $q_{\alpha\beta}$ defined in (14). In the limit $n \to 0$ it becomes a function $q(x)$ which can be derived from (14) and (57)

$$q(x) = \pi^{-d/2} R^{-d}(x) \tag{59}$$

with $R(x)$ defined by (58).

6. Discussion

The replica-symmetry breaking means existence of a numerous equilibrium states separated by high (infinite in thermodynamic limit) energy barriers (Mezard *et al* 1984). The chain configuration in each of these states is fixed up to characteristic scale $R(x_0) \sim R^*$. The physical meaning of it is that each energy valley, or pure state, represents a 'tube' of characteristic diameter R^* and all chain conformations belonging to this valley are within this tube. It is clear that R^* shows the degree of fluctuations in each state and is analogous to Edwards–Anderson parameter in spin-glasses. The character of replica-symmetry breaking (RSB) (i.e. the dependence of the order parameter Q on x) reflects the structure of conformational space, and shows to what extent are valleys corresponding to lowest energies similar. The conventional interpretation of the RSB via the relation $P(q) = \mathrm{d}x/\mathrm{d}q$ (Mezard *et al* 1984) with

$$P(q) = \sum_{a,b} p_a p_b \delta(q - q_{ab}) \tag{60}$$

where summation is taken over all pure states (free energy valleys), q_{ab} is the structural overlap between states. It is clear that $P(q)$ serves as a good 'probe' of the structure of configurational space. It shows whether low-energy valleys (i.e. the ones having high Boltzmann probabilities p_a have similar or different conformations. Using the relation between $P(q)$ and $q(x)$ we have

$$P(q) = q^{(4-5d)/4d} \qquad \text{at} \quad q < q(1). \tag{61}$$

The plot of this function is shown in figure 2. This type of the function $P(q)$ implies that it is possible (and the probability of it is rather high) to find low-energy valleys with some structural similarity ($P(q)$ is non-zero at $q/q(1) \sim 1$). Also continuity of the function $q(x)$ implies that configurational space in this system is ultrametric.

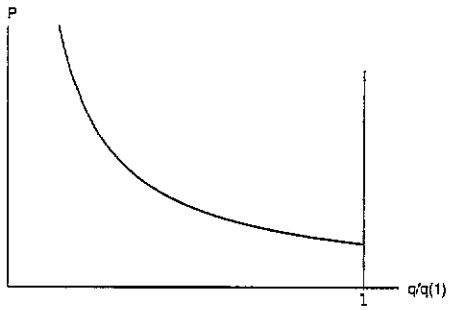

Figure 2. Plot of the function $P(q)$.

As for biological applications of obtained results it should be noted that for $d = 3$ RNA molecules considered as randomly branched trees differ significantly from proteins. As was shown in Shakhnovich and Gutin (1989c) proteins at $d = 3$ can be described by the REM. Replica symmetry breaking and corresponding 'freezing, with the formation of unique structure in linear heteropolymers at $d = 3$ occurs only at some finite heterogeneity (or in other words, at sufficiently low temperature). Space structure in proteins is frozen up to microscopic scale, i.e. micro-state coincides with pure-state. As for RNA molecules some, space structure (but fluctuating on scale R^*) exists at arbitrary heterogeneity (temperature). However this structure is not defined microscopically: at low heterogeneity of a chain it fluctuates strongly and the scale of these fluctuations decrease as heterogeneity increases. Also the fact that low-energy conformations are likely to have some structural similarity in the case of RNA implies that configurational space of such molecules is 'smooth'. Such difference in equilibrium behaviour may lead principally to a different nature of self-organization of proteins and RNA molecules.

Acknowledgments

This work was supported in part by the Petroleum Research Fund administered by ACS through grant No PRF-25536,G6 and by the David and Lucille Packard Fellowship (to ES).

References

Binder K and Young P 1986 *Rev. Mod. Phys.* **68** 801
Bryngelson J and Wolynes P 1987 *Proc. Natl Acad. Sci.* **84** 7524
Daoud M, Pincus P, Stockmayer W H and Witten T 1983 *Macromolecules* **16** 1833
Derrida B 1980 *Phys. Rev. Lett.* **45** 79
Garel T and Orland H 1988 *Europhys. Lett.* **6** 307
Gaunt D S and Flesia S 1991 *J. Phys. A: Math. Gen.* **24** 3655
de Gennes P G 1968 *Biopolymers* **6** 715
Gutell R R, Weiser B, Woesse C R and Noeller H F 1985 *Prog. Nucl. Acid Res. Mol. Biol.* **32** 155–216
Gutin A M, Grosberg A Yu and Shakhnovich E I 1992 *Macromolecules* in press
Higgs P 1992 CEN Saclay SPhT/92-78 *Preprint*
Higgs P and Joanny J F 1991 *J. Chem. Phys.* **94** 1543
Karplus M and Shakhnovich E 1992 *Protein Folding* ed T Creighton (New York: Wiley) ch 4, pp 127–95
Lifshitz I M, Grosberg A Yu and Khokhlov A R 1978 *Rev. Mod. Phys.* **50** 683
Mezard M, Parisi G, Sourlas N, Toulouse G and Virasoro M J 1984 *J. Physique* **45** 843–54
Mezard M and Parisi G 1992a *J. Physique* I **1** 809
—— 1992b *J. Phys. A: Math. Gen.* **25** 4521
Parisi G 1980 *J. Phys. A: Math. Gen.* **13** L155, 1887
—— 1983 *Phys. Rev. Lett.* **50** 1946
Privalov P L and Filimonov V V 1978 *J. Mol. Biol.* **122** 447
Shakhnovich E I and Gutin A M 1989a *Europhys. Lett.* **8** 327
—— 1989b *J. Phys. A: Math. Gen.* **22** 1647
—— 1989c *Biophys. Chem.* **34** 187
—— 1990a *Nature* **346** 773–5
—— 1990b *J. Chem. Phys.* **93** 5967
Sfatos C D, Gutin A M and Shakhnovich E I 1992 *Phys. Rev. Lett.* at press
Stepanow S, Schulz M and Sommer J U 1992 *Europhys. Lett.* **19** 273–7