The lack of long-range correlations is a necessary condition for a functional biologically active protein

E Sh Mamasakhlisov, V F Morozov and M S Shahinian

Yerevan State University, Department of Physics, A Manoogian st 1, 375049 Yerevan, Armenia

Abstract. We studied a random heteropolymer chain with a Gaussian distribution of types of monomers. Long-range correlations between types of monomers were introduced. The mean-field analysis of such a heteropolymer indicates the existence of an infinite energy barrier between the heteropolymer random coil and the frozen states. Thus, the frozen state is kinetically unavailable for the random heteropolymer with power-law correlations in the sequence of the monomer. The relationship between our results and some obtained earlier for the DNA introness sequences are discussed.

1. Introduction

The relationship between the sequence and conformation of a protein macromolecule is one of the great unsolved problems in biophysics. At the present time it is widely believed that functional proteins usually form a single compact three-dimensional structure that corresponds to the global energetic minimum in conformational space. Recently this question has been addressed by the study of random heteropolymers and comparing them with proteins. The fact that even chains with random sequences can have a unique ground state characterized by the frozen path of the polymer-chain backbone was first examined in terms of the random energy model (REM) [1, 2]. Subsequent investigations were carried out on the basis of 'microscopic' Hamiltonians in which the interactions between pairs of monomers were assumed to be random, independently taken from a Gaussian distribution [3], or with polymer sequences explicitly present [4–6]. All these models were shown to exhibit a freezing-phase transition for the random chain.

Recently, Shakhnovich and Gutin [7] found that to have such a minimum it is sufficient that an amino acid sequence forms an uncorrelated random sequence. These results give rise to the question: Is the lack of long-range correlations in protein sequences a *necessary* condition for a three-dimensional biologically functional structure formation? Here we give some positive answers to this question.

From this point of view it is very interesting that recently some results regarding long-range (scale-invariant) correlations in non-coding DNA sequences were obtained [8]. For example it was reported that only non-coding DNA sequences exhibit long-range correlations. This finding, had some support [9, 10]; however, other authors [11, 12] disagreed. For example Voss [10] recently proposed that coding as well as non-coding DNA sequences display long-range power-law correlations in their base-pair sequences.

In this paper, using the above-mentioned problem, we will examine the heteropolymer with the quenched random sequence described by the set of random variables σ_k

characteristic of each monomer. In the past [4–6] the monomer species were considered as independent random variables or, as in [13], were examined using short-range correlations (with exponential decay) between types of monomers. Here we investigate the folding problem for the random heteropolymer with the quenched sequence monomer in the presence of long-range (power-law) correlations.

2. Model and mean-field theory

Let us discuss the heteropolymer chain with a frozen sequence of monomers using a Hamiltonian function of the monomer coordinates $\{r_i\}$. Our model Hamiltonian can be written

$$H = \frac{T}{a^2} \sum_i (r_i - r_{i+1})^2 + \sum_{i<j} B_{ij} \delta(r_i - r_j) + C \sum_{i<j<k} \delta(r_i - r_j)\delta(r_j - r_k) \tag{2.1}$$

where $B_{ij} = B_0 + B\sigma_i\sigma_j$ and C are virial coefficients describing two- and three-particle interactions, σ_k is a variable of the species of kth monomer, a is the statistical segment length and T is the temperature. We work in units where $k_B = 1$.

Earlier reports [1–6] discussed this problem using σ_k statistically independent variables. Here we are using σ_k as random variables with long-range correlations, characterized by the Gaussian distribution function in the form:

$$P\{\sigma\} \propto \exp[-\tfrac{1}{2}(\sigma, \hat{K}^{-1}\sigma)]$$
$$\sigma = (\sigma_1, \dots, \sigma_N) \tag{2.2}$$

where $\hat{K} = \|K_{ij}\|$ is the matrix describing correlations in the chain sequence:

$$K_{ij} = \langle \sigma_i \sigma_j \rangle_P. \tag{2.3}$$

There are several reasons for the existence, in non-coding regions of DNA (introns) only, of scale-invariant long-range correlations in DNA sequences (see, e.g. [14]). The correlation function of the monomer species in this case has the form [14]:

$$K(l) \propto l^{\beta-1} \tag{2.4}$$

where $0 < \beta < 1$ and $l = |i - j|$.

Now we are going to find the free energy F. The standard way to derive the partition function of a system with quenched disorder is to use the replica device:

$$F = \langle F(\sigma) \rangle_P = -T \lim_{n \to 0} \frac{\partial}{\partial n} \langle Z(\sigma)^n \rangle_P \tag{2.5}$$

where $\langle \rangle_P$ means the average over all possible realizations of σ. In these terms the averaged value of the partition function will come to:

$$\langle Z^n \rangle_P = \int \mathrm{D}r_i^\alpha \, g(r_{i+1}^\alpha - r_i^\alpha) \exp\left(-C \sum_\alpha \sum_{ijk} \delta(r_i^\alpha - r_j^\alpha)\delta(r_i^\alpha - r_k^\alpha) \right)$$
$$\times \exp\left(-B_0 \sum_\alpha \sum_{ij} \delta(r_i^\alpha - r_j^\alpha) \right) \left\langle \exp\left(-B \sum_{ij} \sigma_i \sigma_j \sum_\alpha \delta(r_i^\alpha - r_j^\alpha) \right) \right\rangle_P \tag{2.6}$$

where α are replica indices, r_i^a describes the position of the ith monomer of replica α in three-dimensional space and $g(r_i^a - r_j^a)$ is the Gaussian normalized probability distribution

such that $\int \mathrm{d}\boldsymbol{r}\, g(\boldsymbol{r}) = 1$. After linearization over the $\sum_i \sigma_i \delta(\boldsymbol{x} - \boldsymbol{r}_i^\alpha)$ and putting into the expression for $\langle Z^n \rangle_P$ the value of the distribution function $P\{\sigma\}$, we will be led to:

$$
\langle Z^n \rangle_P \propto \int \mathrm{D}\boldsymbol{r}_i^\alpha\, g(\boldsymbol{r}_i^\alpha - \boldsymbol{r}_{i+1}^\alpha) \exp\left(-C \sum_\alpha \int \mathrm{d}\boldsymbol{x}\, \hat{\rho}_\alpha^3(\boldsymbol{x}) - B_0 \sum_\alpha \int \mathrm{d}\boldsymbol{x}\, \hat{\rho}_\alpha^2(\boldsymbol{x}) \right)
$$

$$
\times \int \mathrm{D}\Psi_\alpha(\boldsymbol{x}) \exp\left[\frac{1}{2} \sum_{\alpha < \beta} \int \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \Psi_\alpha(\boldsymbol{x}) \Psi_\beta(\boldsymbol{y}) \hat{q}_{\alpha\beta}(\boldsymbol{x}, \boldsymbol{y}) \right.
$$

$$
\left. - \frac{1}{2|B|} \sum_\alpha \int \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \Psi_\alpha(\boldsymbol{x}) \Psi_\beta(\boldsymbol{y}) \delta(\boldsymbol{x} - \boldsymbol{y}) \right] \tag{2.7}
$$

where

$$
\hat{\rho}_\alpha(\boldsymbol{x}) = \sum_i \delta(\boldsymbol{x} - \boldsymbol{r}_i^\alpha)
$$

$$
\hat{q}_{\alpha\beta}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{ij} K_{ij} \delta(\boldsymbol{x} - \boldsymbol{r}_i^\alpha) \delta(\boldsymbol{x} - \boldsymbol{r}_j^\beta). \tag{2.8}
$$

Thus

$$
\langle Z^n \rangle_P \propto \int \mathrm{D}\rho_\alpha \mathrm{D}q_{\alpha\beta} \exp[-F(\rho_\alpha, q_{\alpha\beta})] \tag{2.9}
$$

where $F(\rho_\alpha, q_{\alpha\beta}) = E(\rho_\alpha, q_{\alpha\beta}) - S(\rho_\alpha, q_{\alpha\beta})$ is the free-energy functional, $E(\rho_\alpha, q_{\alpha\beta})$ is the conformational energy and $S(\rho_\alpha, q_{\alpha\beta})$ is the entropy of n polymer chains which correspond to polymer-chain residue densities $\{\rho_\alpha\}$ and two-replica overlap parameters $\{q_{\alpha\beta}\}$:

$$
-E(\rho_\alpha, q_{\alpha\beta}) = -C \sum_\alpha \int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, (\hat{\rho}_\alpha(\boldsymbol{x}))^3 - B_0 \sum_\alpha \int \mathrm{d}x\, (\hat{\rho}_\alpha^2(x)) + \ln \int \mathrm{D}\Psi_\alpha(\boldsymbol{x})
$$

$$
\times \exp\left\{ \frac{1}{2} \sum_{\alpha\beta} \iint_{xy} \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \Psi_\alpha(\boldsymbol{x}) \Psi_\beta(\boldsymbol{y}) q_{\alpha\beta}(\boldsymbol{x}, \boldsymbol{y}) - \frac{1}{2|B|} \right.
$$

$$
\left. \times \sum_\alpha \iint_{xy} \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \Psi_\alpha(\boldsymbol{x}) \Psi_\beta(\boldsymbol{y}) \delta(\boldsymbol{x} - \boldsymbol{y}) \right\} \exp[S(\rho_\alpha, q_{\alpha\beta})]
$$

$$
= \int \mathrm{D}\boldsymbol{r}_i^\alpha\, g(\boldsymbol{r}_{i+1}^\alpha - \boldsymbol{r}_i^\alpha) \delta(\rho_\alpha - \hat{\rho}_\alpha) \delta(q_{\alpha\beta} - \hat{q}_{\alpha\beta}). \tag{2.10}
$$

In the mean-field approximation we need to minimize the free-energy functional $F(\rho_\alpha, q_{\alpha\beta})$ over the one- and two-replica order parameters $\rho_\alpha, q_{\alpha\beta}$.

The expressions obtained above are identical to those obtained in [4] for the random sequence without correlations. The main difference is in the two-replica overlap parameter $q_{\alpha\beta}$ definition (see equation (2.7)).

Let us make a Fourier transformation of the order parameter of the system:

$$
q_{\alpha\beta}(\boldsymbol{r}) = V^{-1} \sum_{k \neq 0} q_{\alpha\beta}(\boldsymbol{k}) \exp(\mathrm{i}\boldsymbol{k}\boldsymbol{r}) \tag{2.11}
$$

where V indicates the volume used by the macromolecule. This transformation will lead us to a new expression for conformational energy:

$$
-E(q) = \ln \int \mathrm{D}\Psi_\alpha(\boldsymbol{k}) \exp\left\{ -\frac{V}{2} \sum_{\alpha\beta} \sum_{k \neq 0} \Psi_\alpha(\boldsymbol{k}) \Psi_\beta(-\boldsymbol{k}) \left[\frac{\delta_{\alpha\beta}}{|B|} - q_{\alpha\beta}(\boldsymbol{k}) \right] \right\}. \tag{2.12}
$$

Using Gaussian properties this integral expression for conformational energy can be rewritten as:

$$-E(q) = \text{const} - \frac{1}{2} \sum_{k \neq 0} \ln \det P_{\alpha\beta}(\mathbf{k})$$

$$P_{\alpha\beta}(\mathbf{k}) = \frac{\delta_{\alpha\beta}}{|B|} - q_{\alpha\beta}(\mathbf{k}). \tag{2.13}$$

The microscopic order parameter of the system can be displayed in the following form:

$$\hat{q}_{\alpha\beta}(\mathbf{x} - \mathbf{y}) = \hat{Q}_{\alpha\beta}(\mathbf{x} - \mathbf{y}) + \sum_{i \neq j} K_{ij} \delta(\mathbf{x} - \mathbf{r}_i^\alpha) \delta(\mathbf{y} - \mathbf{r}_j^\beta) \tag{2.14}$$

where $\hat{Q}_{\alpha\beta}(\mathbf{x} - \mathbf{y})$ is a two-replica overlap parameter used in papers dedicated to random heteropolymers with non-correlated sequences [3, 4]:

$$\hat{Q}_{\alpha\beta}(\mathbf{x} - \mathbf{y}) = \sum_i \delta(\mathbf{x} - \mathbf{r}_i^\alpha) \delta(\mathbf{y} - \mathbf{r}_i^\beta). \tag{2.15}$$

From the normalization condition

$$\int \hat{Q}_{\alpha\beta}(\mathbf{x} - \mathbf{y}) \, \mathrm{d}\mathbf{x} = \rho_\alpha \tag{2.16}$$

the order parameter $\hat{Q}_{\alpha\beta}(\mathbf{x} - \mathbf{y})$ was found using the method of Shakhnovich and Gutin [3]

$$\hat{Q}_{\alpha\beta}(\mathbf{x} - \mathbf{y}) = \frac{\rho}{R^d} \varphi_{\alpha\beta}\left(\frac{\mathbf{x} - \mathbf{y}}{R}\right) \tag{2.17}$$

where R is the characteristic scale of the two-replica overlap and

$$\int \mathrm{d}\mathbf{z} \, \varphi_{\alpha\beta}(\mathbf{z}) = 1. \tag{2.18}$$

The order parameter of the system $q_{\alpha\beta}$, obviously satisfies to normalization conditions as follows:

$$\int \mathrm{d}\mathbf{y} \, q_{\alpha\beta}(\mathbf{x}, \mathbf{y}) = \bar{K} \rho_\alpha(\mathbf{x})$$

$$\bar{K} \equiv \sum_j K_{ij}. \tag{2.19}$$

Because the thermal average of quantity $\delta(\mathbf{x} - \mathbf{r}_i^\alpha) \delta(\mathbf{y} - \mathbf{r}_j^\beta)$ may be interpreted as a prior probability of corresponding localization of the α replica ith residue and the β replica jth residue (Prob $= P_{ij}^{\alpha\beta}(x, y)$), the order parameter $q_{\alpha\beta}(k)$ required may be found in the form

$$q_{\alpha\beta}(\mathbf{k}) = (K\rho/R^d) \varphi_{\alpha\beta}\left(\frac{\mathbf{x} - \mathbf{y}}{R}\right) + \sum_{i \neq j} K_{ij} P_{ij}^{\alpha\beta}(\mathbf{x}, \mathbf{y}). \tag{2.20}$$

Taking into account the system translation invariance probability distribution $P_{ij}^{\alpha\beta}(\mathbf{x}, \mathbf{y})$ we find

$$P_{ij}^{\alpha\beta}(x, y) = \text{constant } P_{ij}^{\alpha\beta}(y|x) \tag{2.21}$$

where $P_{ij}^{\alpha\beta}(y|x) = \text{Prob}\{\beta$ replica jth residue situated in point y, if the α replica ith residue situated in point $x\}$ is the conditional probability distribution.

It is known [15] that the polymer chain behaviour in the globular state is described by Gaussian statistics. Accordingly, by analogy with [3,4] and taking into account the normalization condition (2.19), we will use the following equation for $q_{\alpha\beta}$:

$$q_{\alpha\beta}(\boldsymbol{x} - \boldsymbol{y}) = K Q_{\alpha\beta}(\boldsymbol{x} - \boldsymbol{y}) + \sum_{i \neq j} K_{ij} P_{ij}^{\alpha\beta}(\boldsymbol{x} - \boldsymbol{y})$$

$$P_{ij}^{\alpha\beta}(\boldsymbol{x} - \boldsymbol{y}) = \int \mathrm{d}z \, P(\boldsymbol{z} - \boldsymbol{x}, |j - i|) \frac{\rho}{N R^d} \varphi_{\alpha\beta}\left(\frac{\boldsymbol{z} - \boldsymbol{y}}{R}\right)$$

(2.22)

where

$$P(\boldsymbol{x} - \boldsymbol{z}, |j - i|) \propto (a^2|j - i|)^{-d/2} \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{z})^2}{a^2|j - i|}\right).$$

(2.23)

Here $P(r, |j - i|)$ is the probability distribution of the end-to-end vector r for the Gaussian polymer chain. After simplifying in the limit $N \to \infty$ we will come to a new equation for Fourier transformation of the order parameter:

$$\tilde{q}_{\alpha\beta}(\boldsymbol{k}) = \frac{\rho}{a^d} \tilde{\varphi}_{\alpha\beta}(\boldsymbol{k}R) \frac{1}{N} \sum_i \sum_{j \neq i} K(|j - i|) \exp\{-a^2 k^2 |j - i|\}$$

$$\cong K\rho\tilde{\varphi}_{\alpha\beta}(\boldsymbol{k}R) + 2\rho \sum_{l \geqslant 1}(l) \exp(-a^2 k^2 l)\tilde{\varphi}_{\alpha\beta}(\boldsymbol{k}R) \equiv \rho\tilde{\varphi}_{\alpha\beta}(\boldsymbol{k}R)A(\boldsymbol{k})$$

(2.24)

where $k = |\boldsymbol{k}|$. It is obvious that without correlations in the polymer-chain sequence our results reduce to those obtained recently in [4].

Using the results of [16, 4] leads us to the following form of conformational energy, in limit $n \to 0$:

$$\frac{E}{n} \cong -\frac{1}{2} \sum_k \int_0^1 \frac{\mathrm{d}x}{x^2} \ln \lambda_k(x)$$

(2.25)

where

$$\lambda_k(x) = 1/|B| - \rho A(k) - \int_x^1 \mathrm{d}y \, M_k(y) - x M_k(x)$$

(2.26)

and the Parisi function $M_k(x)$ parametrizing the off-diagonal elements of the hierarchical matrix $P_{\alpha\beta}(k)$ in the $n \to 0$ limit.

Now we have to minimize the free-energy functional (see equations (2.8) and (2.9)). It is known [3,4], that the replica-symmetric solution is invalid for random heteropolymers, because in this case the entropy contribution has the form [3]:

$$S\{\rho_\alpha, q_{\alpha\beta}\} = (n - 1)\Delta S(R)$$

(2.27)

where $\Delta S(R)$ is the loss of entropy of the ideal polymer chain constrained in a tube of diameter R; here

$$\Delta S(R) \cong \begin{cases} -Na^2/R^2 \, (R \gg a) \\ Nd\ln(R/a) \, (R \ll a) \end{cases}.$$

(2.28)

Following the Parisi ansatz for one-step replica symmetry breaking (RSB), for n replicas there are n/x groups with x replicas per group. The entropy loss is therefore:

$$S\{\rho_\alpha, q_{\alpha\beta}\} = \frac{n}{x}(x - 1)\Delta S(R).$$

(2.29)

In the same one-step RSB the energy contribution (for $n \to 0$) becomes

$$E = \frac{E\{q_{\alpha\beta}\}}{n} \cong \frac{T}{2} \int \frac{\mathrm{d}\boldsymbol{k}}{R^d} \left\{ \frac{1}{x} \ln[b - x\rho A(\boldsymbol{k}/R)(1 - \tilde{\varphi}(\boldsymbol{k}))] + \left(1 - \frac{1}{x}\right) \ln[b] \right\}$$

(2.30)

where $b - 1/|B|$.

For the subsequent minimization consider the function $A(k)$ behaviour (see equation (2.23)) defined above

$$A\left(\frac{k}{R}\right) = K + 2\sum_{l\geqslant 1} K(l)\exp\left(-l\frac{a^2 k^2}{R^2}\right).$$

(2.31)

For the correlation function (2.4) the $A(k/R)$ can be evaluated as

$$A\left(\frac{k}{R}\right) \cong K\left\{1 + 2\Gamma(\beta)\left(\frac{ak}{R}\right)^{-2\beta}\right\} \qquad \text{for } \frac{ak}{R} \ll 1.$$

(2.32)

Consequently, the function $A(k/R)$ is increased monotonically with R/k and thus for any k value we can find the R scale such that the energy contribution in equation (2.30) will diverge as $\ln(b - \text{constant } (R/a)^{2\beta})$. From the form of equations (2.30) we can see that there is a wide-scale R region, which is forbidden energetically, as the dependence E versus R has the form shown by figure 1(a). The entropy loss $\Delta S(R)$ due by restrictions with scale R has the form shown by figure 1(b). It is obvious that, for the small enough values of R-scale, we have a situation identical with that of [3, 4] (for the random heteropolymer (RHP) with a non-correlated sequence). Consequently, free energy for the one-step RSB has one maximal value at the $R \approx 0$ (see figure 1(c)) and, correspondingly, one thermodynamically stable state. Moreover, the free energy for the one-step RSB has the infinite energy barrier due to divergence (see above), which separates the stable state with $R \approx 0$ from the other (unfrozen) states. In the case of the two-step RSB we carried out the calculations analogous to equation (2.30). For this scheme the conformational energy also diverges at large enough replica overlap scales. Thus, the above results are not due to the one-step RSB approximation; however, they reflect the system properties examined here.
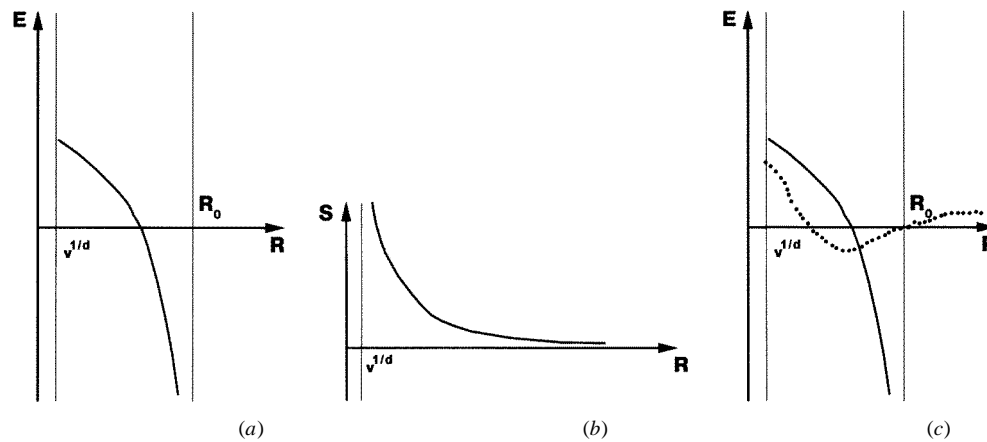

Figure 1. (*a*) The conformational energy dependence versus the scale of replicas overlap (R). Here ν is the excluded volume of chain residue and R_0 is the scale of conformational energy divergence. (*b*) The conformational entropy plotted versus the scale of replicas overlap (R). Here ν is the excluded volume of the chain residue and R_0 is the scale of the conformational energy divergence. (*c*) The full curve is the free-energy dependence versus the scale of replicas overlap (R). The broken curve free-energy plotted versus scale R for the RHP with a non-correlated sequence of residues (see, e.g. [3, 4]).

3. Discussion

As well as other investigated peculiarities, there appear to be long-range correlations of residue sequences in the heteropolymer chain folding. The interresidual correlations are defined by the Gaussian distribution function (2.2) with a non-diagonal correlation matrix (2.3). The probability distribution does not factor on the terms corresponding to the chain residues which are different from the non-correlated sequence case. In this paper the standard techniques, developed in [3] were used. The results obtained above are mathematically similar to those in [4], but the two-replica order parameter redefinition led to unexpected physical properties in the system under examination.

In the case of the power-law correlations decay (see equation (2.4)), the off-diagonal terms which have contributions in the order parameter $q_{\alpha\beta}$ (see equation (2.8)) led to the qualitatively different behaviour of the energetical term (2.30). Taking into account estimation (2.33) we can see that each harmonic k energetic contribution characterized by some space scale R of divergence, is defined by the expression

$$b - x\rho A \left(\frac{k}{R}\right)(1 - \tilde{\varphi}(k)) = 0. \tag{3.1}$$

At the large enough values of R the last expression may be rewritten as

$$b \cong 2\rho x(1 - \tilde{\varphi}(k))\Gamma(\beta)\left(\frac{R}{ak}\right)^{2\beta}. \tag{3.2}$$

Expressions (2.17) and (2.18) show that the characteristic scale of the $\varphi(z)$ function decay is $|z| \approx 1$ and $\varphi(z) = \varphi(|z|)$ has the form schematically represented in figure 2. Consequently, there exists k^*, such that for any $k > k^*$ $\tilde{\varphi}(k) < 1$. So, for example, for $\varphi(r) \propto \exp(-r^2/2)$ we have $\tilde{\varphi}(k) \propto \exp(-k^2/2)$. Thus, for $k > k^*$ equation (3.5) has the solution, defined by
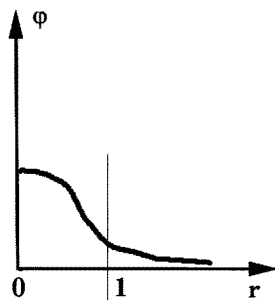

Figure 2. The overlap function behaviour plotted versus the dimensionless scale of replicas overlap.
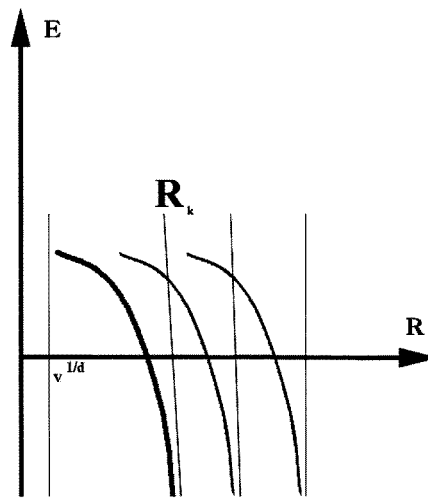
Figure 3. The scheme of different harmonics contributions for the conformational energy. R_k is the scale of divergence for the wavevector length k (see equation (3.3)).

the expression

$$R \cong ak \left(\frac{b}{2\rho x(1 - \tilde{\varphi}(k))} \right)^{\frac{1}{2}\beta} \tag{3.3}$$

and the system energy may be considered as a superposition of different harmonics contributions, (see figure 3). It is obvious that for large enough values of R-scale the system energy becomes divergent. Thus, for the polymer globule the only frozen state (with $R \approx 0$) is stable.

Taking into account the normalization condition for $g(r)$ (see the explanations of equation (2.6)), the random-coil state-free energy in the polymer globule mean-field theory [15] evaluated as $F = 0$. Consequently, our system has two stable states. One, corresponding to the frozen-chain backbone path and the other a random-coil state. Because in the large enough values of R-scale the system energy became divergent these stable states are separated by an infinite energetic barrier. This result may be interpreted as follows. The heteropolymer chain with a long-correlated sequence can exist in a folded $R \approx 0$ or random-coil state, but the folded state is kinetically not available. Thus, the power-law correlations led to the random heteropolymer folding impossibility.

It is of interest that the correlations exponential decay

$$K(l) \propto \exp(-l/\xi) \tag{3.4}$$

in the limit $\xi \gg 1$ is equivalent to $\beta = 1$ in (2.4), which is the marginal case of maximal long-range correlations in the sequence

$$
\begin{aligned}
\tilde{q}_{\alpha\beta}(k) \quad &= \rho \tilde{\varphi}_{\alpha\beta}(kR) \left\{ 1 + 2 \sum_{l \geqslant 1} \exp[-l(a^2 k^2 + 1/\xi)] \right\} \\
&\cong \rho \tilde{\varphi}_{\alpha\beta}(kR) \left\{ 1 + 2 \sum_{l \geqslant 1} \exp(-a^2 k^2 l) \right\}.
\end{aligned}
\tag{3.5}
$$

If we have $\xi \approx 1$, then

$$\tilde{q}_{\alpha\beta}(k) \cong \rho \tilde{\varphi}_{\alpha\beta}(kR) \left\{ 1 + 2 \sum_{l \geqslant 1} \exp(-l/\xi) \right\} \tag{3.6}$$

which is completely equivalent to the order parameter obtained in [4] for the heteropolymer with a non-correlated sequence. It is quite natural because, in the case of $\xi \approx 1$, the chain sequence may be divided into enough small pieces that will be statistically independent.

The above results are in agreement only with the hypothesis of long-range correlations in non-coding DNA sequences [8]. Recently, Shakhnovich and Gutin [7] found that for RHP to have the energetically stable folded state it is sufficient that the sequence of monomers forms an uncorrelated random sequence. Our results show that the lack of long-range correlation sequences is a necessary condition for RHP folding possibility.

References

[1] Derrida B 1980 *Phys. Rev. Lett.* **49** 79
[2] Bryngelson J B and Wolynes P G 1987 *Proc. Natl Acad. Sci., USA* **84** 7524
[3] Shakhnovich E and Gutin A 1989 *Biophys. Chem.* **34** 187
[4] Sfatos C, Gutin A and Shakhnovich E 1993 *Phys. Rev.* E **48** 465
[5] Garel T and Orland H 1988 *Europhys. Lett.* **6** 597
[6] Sfatos C, Gutin A and Shakhnovich E 1994 *Phys. Rev.* E **50** 2898
[7] Shakhnovich E and Gutin A 1990 *Nature* **346** 773

 [8] Peng C-K, Buldyrev S V, Goldberger A L, Havlin S, Sciortino F, Simons M and Stanley H E 1992 *Nature* **356** 168
 [9] Li W and Kanelko K 1992 *Europhys. Lett.* **17** 655
[10] Woss R 1992 *Phys. Rev. Lett.* **68** 3805
[11] Nee S 1992 *Nature* **357** 450
[12] Prabhu V V and Claverie J-M 1992 *Nature* **357** 782
[13] Gutin A, Sfatos C and Shakhnovich E 1994 *J. Phys. A: Math. Gen.* **27** 7957
[14] Stanley H, Buldyrev S V, Goldberger A L, Goldberger Z G, Havlin S, Mantegna R N, Ossadnik S M, Peng C-K and Simons M 1994 *Physica* **205A** 214
[15] Grossberg A Yu and Khokhlov A R 1994 *Statistical Physics of Macromolecules* (New York: AIP)
[16] Mezard M and Parisi G 1991 *J. Physique* I **1** 809