===== **MOLECULAR BIOPHYSICS** =====

Hidden Symmetry Effects in the Dynamics of Linear Polymers and Biopolymers

K. V. Shaitan[a, b, *]

[a] *Moscow State University, Moscow, 119991 Russia*
[b] *Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, Moscow, 119991 Russia*
*e-mail: shaytan49@yandex.ru

Abstract—The effects of two types of symmetry for linear polymers are considered: spatial symmetry with respect to translations and rotation of the macromolecule as a whole in a homogeneous viscous medium, and symmetry with respect to permutations of identical monomer units (or sections of the polymer chain with the same chemical structure) in the chain. It is shown that, in a homogeneous viscous medium, for a macromolecule with rigid bonds during relaxation folding, the conservation law is obeyed—the sum of the rotation velocity vectors around torsion angles is equal to zero. Symmetry with respect to permutations of identical monomer units in the polymer chain under certain conditions leads to the formation of energy funnels with the minimal frustration which correspond to helical or multihelical 3D-structures of polymer chain. In this case, there exist the deepest central funnel and less deep satellite funnels, which all are separated from each other by energy barriers. This topography of the energy funnel corresponds to a number of effects that are observed in the kinetics of protein folding (a volcano-like profile of the free energy surface, the sensitivity of protein refolding to denaturing steps, etc.). When calculating the topography of the free energy surface, the characteristic temperature parameter T_0 arises and this is defined as potential energy gain due to chain folding per one conformational degree of freedom. It is shown that at $T > 0.26\ T_0$, the spatial structure of the folded polymer is destroyed. The parameter T_0 and the denaturation temperature in the system under consideration arise from the basic mathematical principles of the arrangement of energy landscapes in the configuration space of torsion angles with the topology of a multidimensional torus and symmetry considerations regarding the permutation of identical monomer units. At biopolymer denaturation temperatures, for example, on the order of 60°C, this ratio leads to an estimate of the energy of nonvalent bonds of monomers on the order of 2.5 kcal/mol per one conformational degree of freedom, which is very similar to hydrogen bonds in aqueous medium. The two types of symmetries under study and their influence on the dynamics of macromolecules and the topography of the energy landscapes of linear polymers can occur simultaneously under certain conditions. The effects that come with it may provide additional information about the prebiological physicochemical evolution of macromolecules in association with the formation of a pool of linear polymers with unique spatial structures.

Keywords: regularities in the dynamics of linear polymers in a viscous medium, energy landscapes of linear polymers and biopolymers, spatial structure of biopolymers, folding dynamics, minimum frustration principle for an energy funnel, problems of calibration of force fields, multidimensional Fourier series, free energy of biopolymers, critical temperature of denaturation, physicochemical molecular evolution

DOI: 10.1134/S0006350922030204

The formation of the spatial structure and dynamics of polymers and biopolymers have been the focus of attention of researchers for many years and have been studied using various experimental and theoretical methods (see, for example, [1–13]). Progress in understanding the fundamental principles underlying the formation of spatial structure and dynamic behavior of biopolymers, is currently limited to rather simple phenomenological concepts (see, for example, [3–5, 7–10, 12]) and the results of computer molecular

modeling of individual biopolymer structures using various force fields and protocols [15–18]. A huge amount of numerical experiments have been accumulated on modeling various biopolymers. However a critical understanding of these in terms of the fundamental physical laws underlying the structural and dynamic organization of biopolymer systems is impossible without further development of general theoretical ideas and corresponding mathematical tools. It should also be noted that, until now, there is no physically meaningful statement of the problem for a number of fundamental questions underlying the origin and functioning of living systems. For example, the

---

Abbreviations: PES, potential energy surface; IPES, ideal potential energy surface; FES, free energy surface.

physical principles underlying the formation of strictly defined (unique) spatial structures of linear polymers (in particular, biopolymers) are not very clear. For instance, are the observed spatial structures of proteins a game of chance, when certain amino acid sequences have a single global minimum of free energy and the possibility of reaching it in a reasonable time, or are there physical patterns which determine and regulate this phenomenon? Approximately 30 years ago, a hypothesis (or the principle) of minimal frustration of the energy funnel [14] arose as applied to the topography of the multidimensional energy surface of polypeptides that form spatial protein structures. However, until now, there were no physical or mathematical arguments in favor of this smooth arrangement of the energy surface (in particular, proteins). Moreover, in the case of a completely random formation of macromolecules with smooth energy funnel that can form unique spatial structures, the problem of their origin in inanimate nature is absolutely incomprehensible. The ambiguity in the most fundamental issues concerning the physical mechanisms of the origin of biopolymer prototypes at the prebiological stage of evolution and in the physical foundations for the formation of unique spatial structures of conformationally labile biomacromolecules in living systems makes us look for and analyze the most basic principles that can be used by Nature to solve these structural problems.

Earlier in a series of papers, we discussed the patterns caused by the influence of the viscosity of a medium on the dynamics of linear polymers [19], as well as the role of the topology of the configuration space of linear polymers in the formation of the potential energy surface and the free energy surface of macromolecules [20−22]. Below, we consider these problems from the point of view of the symmetry of a polymer (biopolymer) with respect to translations and rotations of the macromolecule as a whole in a homogeneous viscous medium, and symmetry with respect to permutations of identical links in a linear polymer chain. Various kinds of symmetries are among the most fundamental characteristics for understanding the properties and physical structure of different objects [23]. As far as we know, the symmetry effects discussed in the article have not been considered before. As we shall see, the both symmetries despite being so different in nature but under certain conditions can act together and turn out to be an important factor in molecular physicochemical evolution towards the formation of linear polymers with a unique spatial structures.

EFFECTS OF SYMMETRY REGARDING TRANSLATIONS AND ROTATIONS OF MACROMOLECULES AS A WHOLE IN A HOMOGENEOUS VISCOUS MEDIUM

It is known that a system of equations of motion of atoms in a linear polymer chain (as well as practically any system of interacting material points) in a homogeneous viscous medium can be written in the following form [24]:

$$\dot{p}_i + \Gamma_i p_i = F_i, \qquad (1)$$

where p_i is the component of the momentum of the ith atom in the chain, Γ_i is the decay equal to the ratio of the coefficient γ_i of the viscous friction of the ith atom to its mass m_i, and F_i is the component of the sum of the forces that act on the ith atom. A dot over a variable means taking the total derivative with respect to time. The solution of Eq. (1) can be represented in the general form

$$p_i(t) = p_i(0)e^{-\Gamma_i t} + e^{-\Gamma_i t}\int_0^t F_i(\tau)e^{\Gamma_i t}d\tau$$

$$= \frac{1}{\Gamma_i}F_i(t) + \left[p_i(0) - \frac{1}{\Gamma_i}F_i(0)\right]e^{-\Gamma_i t} - \frac{1}{\Gamma_i}e^{-\Gamma_i t}\int_0^t \dot{F}_i(\tau)e^{\Gamma_i t}d\tau. \qquad (2)$$

The dependence of the force components on time in Eq. (2) should be understood in such a way that, in the function F_i, which determines the dependence of forces on the coordinates of atoms, we substitute real trajectories or dependences of coordinates on time from the solution of the system of equations (1). The time derivative of the force under the integral on the right side of Eq. (2) is on the order of the ratio of the magnitude of the force to the characteristic time of change in the force or the characteristic time of change in the interatomic distances τ. For conformational motions in an aqueous medium, this time lies in the nanosecond and longer time ranges [2−8]. The decay time $1/\Gamma_i$, as can be seen from the first term in Eq. (2), is the characteristic relaxation time of the particle velocity, which in a liquid (water) lies in the subpicosecond range [25]. Therefore, the order of magnitude of the ratio of the last and first terms in the second part of Eq. (2) will be $1/(\tau\Gamma_i) \sim 10^{-4}$. Thus, at times $t > 1/\Gamma_i \sim 10^{-13}$ s, the solution of the system of equations (1) with high accuracy becomes equivalent to the solution of this system of equations without the first (inertial) terms. It should be noted that, over a time of about $1/\Gamma_i \sim 10^{-13}$ s, the initial conformation of the macromolecule remains virtually unchanged. With an accuracy up to about $1/(\tau\Gamma_i)$ at times $t \gg 1/\Gamma_i \sim 10^{-13}$ s, we obtain the equations of motion in the usual form:

$$\gamma_i \dot{\mathbf{r}}_i = -\frac{\partial U}{\partial \mathbf{r}_i}, \qquad (3)$$

where U is the potential energy, depending on the distances between the particles and their interaction with the environment, and $\mathbf{r}_i$ are the radius-vectors of the positions of particles in space. System of equations of motion (3) in highly viscous media is usually used without specifying the asymptotic conditions for its applicability. However, these conditions must be taken into account when deriving various relations using

solutions (3), seeing that the formulas obtained in this way work with an accuracy of about $1/(\tau \Gamma_i)$.

Earlier, we discussed the effects associated with the addition of a stochastic force (thermal noise) and rigid constraints to equations of type (3) [19]. These additional factors do not fundamentally affect the main conclusions, and in this paper we will not complicate the text below.

If the medium in which the macromolecule is located is homogeneous, then the potential energy and the system of equations (3) are invariant with respect to the displacement of the entire system as a whole by the vector $\mathbf{d}$. Assuming vector $\mathbf{d}$ is small, we write change in potential energy during translation of the system as a whole

$$\delta U = \mathbf{d}\sum_i \frac{\partial U}{\partial \mathbf{r}_i} = 0,$$

or

$$\sum_i \frac{\partial U}{\partial \mathbf{r}_i} = 0. \tag{4}$$

Equation (4) reflects the fact that the sum of all forces inside the macromolecule is equal to 0. Therefore, from the invariance of the system with respect to translations as a whole in a homogeneous viscous medium, we obtain the conservation law

$$\sum_i \gamma_i \dot{\mathbf{r}}_i = 0, \tag{5}$$

or

$$\sum_i \gamma_i \mathbf{r}_i = c\mathbf{R} = \text{const}, \tag{6}$$

where c is some number of dimensions of the coefficient of friction. Its physical meaning can be determined from the following considerations. Let us apply a constant external force $\mathbf{f}_i$ to each particle. In this case the right side of (5) will contain the sum of external forces $\mathbf{f}$ applied to the system of particles, and the equation for the motion of the entire system in a viscous medium will take the form

$$c\dot{\mathbf{R}} = \mathbf{f}. \tag{7}$$

It is clear from Eq. (7) that the coefficient c is the total coefficient of friction:

$$c = \sum_i \gamma_i. \tag{8}$$

In other words, during conformational motions in the absence of external forces, the point $\mathbf{R}$ (conditionally, the dynamic center of the molecule) does not shift. The position of this point is determined by the initial configuration of the system. Physically, this can be understood as the impossibility of giving the molecule a directed motion only due to the forces of viscous friction. If all coefficients of friction are the same, then this point has the same meaning as the center of mass of a system of particles with the same mass. In the case

of the equal values $\Gamma_i = \Gamma$ from the system of Eqs. (1) we also obtain an equation for the change in the total momentum of a macromolecule in a viscous medium:

$$\mathbf{P}(t) = \sum_i \mathbf{p}_i(t) = \mathbf{P}(0)e^{-\Gamma t}. \tag{9}$$

In other words, at times significantly longer than the particle velocity relaxation time, the translational motion of the molecule as a whole stops.

Next, we vectorially multiply the right and left parts of Eqs. (3) by the corresponding radius vectors and sum the equations:

$$\sum_i \gamma_i [\dot{\mathbf{r}}_i \times \mathbf{r}_i] = -\sum_i \left[\frac{\partial U}{\partial \mathbf{r}_i} \times \dot{\mathbf{r}}_i \right]. \tag{10}$$

Taking into account that in a homogeneous medium the potential energy is invariant with respect to rotations of the molecule as a whole, and introducing a rotation vector around some axis $\delta\boldsymbol{\phi}$ so that the direction of the vector coincides with the direction of the axis and the length of the vector is equal to the angle of rotation [24], we obtain the change in the radius-vectors of particles for small angles:

$$\delta\mathbf{r} = [\delta\boldsymbol{\phi} \times \mathbf{r}_i]. \tag{11}$$

Then, for the change in the potential energy during rotation, we obtain:

$$\begin{aligned}\delta U &= \sum \frac{\partial U}{\partial \mathbf{r}_i}\delta\mathbf{r}_i = \sum_i \frac{\partial U}{\partial \mathbf{r}_i}[\delta\boldsymbol{\phi} \times \mathbf{r}_i] \\ &= -\delta\boldsymbol{\phi}\sum_i \left[\frac{\partial U}{\partial \mathbf{r}_i} \times \mathbf{r}_i \right] = 0,\end{aligned} \tag{12}$$

where we used the rule of cyclic permutation of vectors in a mixed product. Since the vector $\delta\boldsymbol{\phi}$ is arbitrary, we obtain from Eqs. (10) and (12) that the sum of friction force moments (as well as the sum of internal force moments in a macromolecule) is equal to 0:

$$\sum_i \gamma_i [\dot{\mathbf{r}}_i \times \mathbf{r}_i] = 0. \tag{13}$$

Physically, this means that, due to the forces of friction, it is impossible to twist the molecule as a whole. In the simplest case of equal values Γ_i, from the system of Eqs. (1) we obtain for the total angular momentum of the system:

$$\mathbf{M}(t) = \sum_i [\mathbf{p}_i(t) \times \mathbf{r}_i(t)] = \mathbf{M}(0)e^{-\Gamma t}, \tag{14}$$

that is, for a time longer than the particle velocity relaxation time, the rotation of the molecule as a whole stops.

Relations (5) and (13) are quite obvious; however, they can be used to obtain another interesting integral of motion for a linear polymer chain with rigid bonds in a viscous medium.

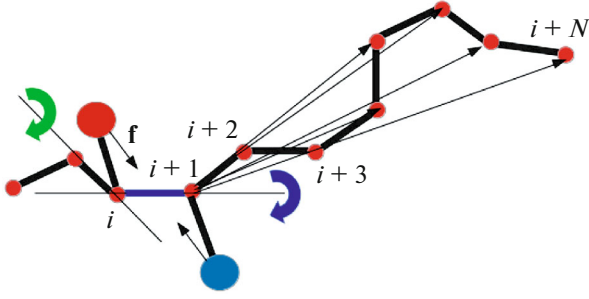We consider a linear polymer chain (Fig. 1), in which the lengths of valence bonds and valence angles

Fig. 1. Rotations about the bonds in the polymer chain as a result of the attractive force **f** of the side groups. In a viscous medium, at a given angular velocity of rotation about the bond between nodes i and $i + 1$, the friction forces will increase as the distance between the chain nodes and the axis of rotation increases. Thus, the friction forces in the chain nodes will also cause rotations about other bonds.

have almost no change during motion. Note that the addition of rigid bonds, which fix the geometric parameters of chemical bonds, to Eq. (1) does not affect the sums of relations (5) or (13). At fixed bond lengths and bond angles, the total shift of an arbitrary ith node occurs in accordance with Eq. (3) only due to rotations around the bonds between the nodes of the chain.

We introduce the vectors of angular velocities of rotation about the torsion angle ϕ_i, which are directed along the link from the ith node of the chain to the $(i + 1)$th node:

$$\dot{\boldsymbol{\phi}}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{|\mathbf{r}_{i+1} - \mathbf{r}_i|} \frac{d\phi_i}{dt}.$$

The linear speed of motion of the ith node is determined as the sum of contributions from turns around all bounds in accordance with Eq. (11):

$$\dot{\mathbf{r}}_i = \sum_{k=1}^{N} [\dot{\boldsymbol{\phi}}_k \times (\mathbf{r}_i - \mathbf{r}_k)] \tag{15}$$

Note that, when considering the instantaneous values of velocities, we do not have problems related to the noncommutativity of rotations around bonds. Rotations around the first and last bonds in the case of a linear chain are equivalent to rotations of the macromolecule as a whole around the corresponding axes. With the exception of terminal atoms, at least two terms in sum (15) associated with rotations around two bonds adjacent to the ith node are equal to 0.

Let us then use the invariance of the system with respect to translations in a homogeneous viscous medium. Substituting expression (15) into relation (5), we obtain

$$\sum_i \gamma_i \dot{\mathbf{r}}_i = \sum_i \gamma_i \sum_k \{[\dot{\boldsymbol{\phi}}_k \times \mathbf{r}_i] - [\dot{\boldsymbol{\phi}}_k \times \mathbf{r}_i]\}$$
$$= \sum_k [\dot{\boldsymbol{\phi}}_k \times c\mathbf{R}] - \left(\sum_i \gamma_i\right) \sum_k [\dot{\boldsymbol{\phi}}_k \times \mathbf{r}_i] = 0. \tag{16}$$

Introducing the vector of the sum of the vectors of the angular velocities of turns around the bonds

$$\dot{\boldsymbol{\Phi}} = \sum_k \dot{\boldsymbol{\phi}}_k, \tag{17}$$

in accordance with expressions (8) and (16) we obtain:

$$[\dot{\boldsymbol{\Phi}} \times \mathbf{R}] = \sum_k [\dot{\boldsymbol{\phi}}_k \times \mathbf{r}_k]. \tag{18}$$

Note that Eq. (18) does not depend on the selection of the origin.

Then we use the isotropy of a homogeneous viscous medium and substitute expression (15) into relations (13). Taking into account the properties of the double-vector product and relation (18), as well as Eqs. (6) and (8), we obtain:

$$\sum_i \gamma_i \sum_k \{[\dot{\boldsymbol{\phi}}_k \times \mathbf{r}_i] \times \mathbf{r}_i - [\dot{\boldsymbol{\phi}}_k \times \mathbf{r}_k] \times \mathbf{r}_i\}$$
$$= \sum_i \gamma_i \sum_k [[\dot{\boldsymbol{\Phi}} \times \mathbf{r}_i] \times \mathbf{r}_i] - \sum_i \gamma_i [[\dot{\boldsymbol{\Phi}} \times \mathbf{R}] \times \mathbf{R}]$$
$$= \sum_i \gamma_i (\dot{\boldsymbol{\Phi}} \cdot \mathbf{r}_i)\mathbf{r}_i - \dot{\boldsymbol{\Phi}} \sum_i \gamma_i \mathbf{r}_i^2 \tag{19}$$
$$+ \sum_i \gamma_i \{\dot{\boldsymbol{\Phi}} \mathbf{R}^2 - \mathbf{R}(\dot{\boldsymbol{\Phi}} \cdot \mathbf{R})\} = 0.$$

Separating the terms with the sum of the angular velocity vectors, we obtain

$$\dot{\boldsymbol{\Phi}} \left(\sum_i \gamma_i \mathbf{r}_i^2 - \sum_i \gamma_i \mathbf{R}^2 \right)$$
$$= \sum_i \gamma_i (\dot{\boldsymbol{\Phi}} \cdot \mathbf{r}_i)\mathbf{r}_i - \sum_i \gamma_i (\dot{\boldsymbol{\Phi}} \cdot \mathbf{R})\mathbf{R}. \tag{20}$$

Note that relation (20), as well as relation (18), does not depend on the selection of the origin of the reference frame $\mathbf{R}_0$:

$$\dot{\boldsymbol{\Phi}} \left(\sum_i \gamma_i (\mathbf{r}_i - \mathbf{R}_0)^2 - \sum_i \gamma_i (\mathbf{R} - \mathbf{R}_0)^2 \right)$$
$$= \sum_i \gamma_i (\dot{\boldsymbol{\Phi}} \cdot (\mathbf{r}_i - \mathbf{R}_0)_i)(\mathbf{r}_i - \mathbf{R}_0) \tag{21}$$
$$- \sum_i \gamma_i (\dot{\boldsymbol{\Phi}} \cdot (\mathbf{R} - \mathbf{R}_0))(\mathbf{R} - \mathbf{R}_0).$$

Hence, selecting a fixed point with the vector $\mathbf{R}$ (6) as the origin, we obtain:

$$\dot{\boldsymbol{\Phi}} \sum_i \gamma_i (\mathbf{r}_i - \mathbf{R})^2 - \sum_i \gamma_i (\dot{\boldsymbol{\Phi}} \cdot (\mathbf{r}_i - \mathbf{R}))(\mathbf{r}_i - \mathbf{R}) = 0. \tag{22}$$

Assuming the total angular velocity vector is not equal to 0 and scalarly multiplying Eq. (22) by it, we come to the relation

$$\dot{\boldsymbol{\Phi}}^2 \sum_i \gamma_i (\mathbf{r}_i - \mathbf{R})^2 \sin^2 \alpha_i = 0, \tag{23}$$

where α_i are the angles between the total angular velocity vector and the particle radius vectors relative to the origin (points $\mathbf{R}$). The factor at the square of the total angular velocity can be equal to 0 only in the exceptional case of a strictly linear configuration of the

macromolecule. In any other case, we obtain a contradiction. Therefore, the sum of the angular velocity vectors rotation around the bonds in a linear polymer in a viscous medium must be equal to 0:

$$\dot{\mathbf{\Phi}} = \sum_k \dot{\mathbf{\phi}}_k = \sum_k \dot{\phi}_k \mathbf{b}_k / b_k = 0. \qquad (24)$$

For clarity, in Eq. (24) we also introduced the values of the angular velocities and the vector $\mathbf{b}_k$ along the bond between neighboring nodes of the chain. In reality, the fulfillment of condition (24) means that, during the relaxation folding of the polymer chain, the process goes in such a way that the directions of rotations around the bonds alternate in some way so as to compensate each other in total (we intuitively felt this when analyzing the situation in Fig. 1). As a result, there is a tendency to form helical or helix-like structures. We observed this effect earlier when modeling the folding dynamics of polypeptide chains [17, 18] and some other linear polymers in a viscous medium [16].

Additional information about the conformational motion characteristics of the chain can be obtained by multiplying (24), for example, scalarly by specially constructed vectors. For example, if we multiply (24) by the $\mathbf{l}$ vector:

$$\mathbf{l} = \sum_i \mathbf{b}_i,$$

which connects the ends of the chain, we obtain

$$\sum_k \dot{\phi}_k l_k = 0; \quad l_k = (\mathbf{b}_k \mathbf{l})/b_k. \qquad (24.1)$$

Note that the discussed patterns of the polymer chain dynamics in a homogeneous viscous medium are most applicable to the conditions in which all chain nodes are exposed to the solution, i.e., in the situation of an unfolded state or a sufficiently loose coil. If the chain is in an unfolded state, then the scalar products l_k in the sum (24.1) are predominantly positive. Therefore, the angular velocities must often change sign in order to fulfill condition (24.1). As the chain undergoes folding, the orientation of the direction of the bond vectors relative to vector $\mathbf{1}$ will change, and the alternation of the signs of the angular velocities will become less pronounced. This situation was observed earlier when modeling the folding of a long polypeptide chain [17].

Note that when fluctuations of the medium are taken into account in the equations of motion and when we proceed from Eq. (1) to the Langevin equations [19, 25], the relations obtained above hold only on average over the ensemble. On the other hand, relations such as (24) can give ideas for performing numerical and conventional experiments for additional elucidation of the mechanisms of self-organization of the spatial structure of macromolecules and their complexes under various conditions. We also note that the combination of the rules of motion of a representative point, obtained here and in [19], during conformational motions in a viscous medium and the features of the topography of the energy landscapes of biopolymers and similar structures [20−22] may be essential for understanding the physical principles of the origin of linear polymers with unique spatial structures (see below).

In the well-known paper [9] dedicated to the 50th anniversary of the folding problem, unresolved issues were noted, among which the following is also relevant in this context: the practical absence of experimental data on the structure of the energy landscape required for folding and the absence of a quantitative microscopic picture of understanding the physics and folding rules for arbitrary sequences. Below we will consider what additionally can be done in these directions proceeding from the topology of the configuration space and symmetry considerations with respect to the permutation of identical monomer links of a linear polymer chain.

EFFECTS OF SYMMETRY REGARDING THE PERMUTATION OF THE SAME UNITS IN A LINEAR POLYMER CHAIN

It is well known [26] that the spatial structure and dynamics of linear polymers are determined by rotations at dihedral (torsion) angles around single chemical bonds with internal rotation barriers on the order of several kcal/mol. Stretching vibrations do not contribute significantly to the change in the spatial configuration of the macromolecule. Bond angle deformation may be of some importance in reducing the barrier height in the case of steric restrictions during rotation around the bonds [11]. For further consideration, this effect is also of no significant importance.

Thus, the potential energy of a linear polymer chain (more precisely, its conformation-dependent part) can be represented as a function of N torsion angles ϕ_i. For a linear polymer, this is a function of a large number of variables, each of which changes cyclically from $-\pi$ to $+\pi$. The area of definition of each angle is the points of the circle. The domain of definition of the set of N angles is a topological product of circles or a multidimensional torus (hypertorus) [27] (Fig. 2). In other words, the potential energy function is definded in configuration space with the topology of a multidimensional torus.

A torus (and a hypertorus) can be unfolded by making cells of the appropriate dimension in Euclidean space with periodic boundary conditions (Fig. 2b). Such a scan may be more convenient for the perception of general geometric relations, and we will use it by default below, paying attention to those moments where the periodicity of the cells plays a significant role.

It is known that the topology of the space on which functions are defined has a significant effect on the
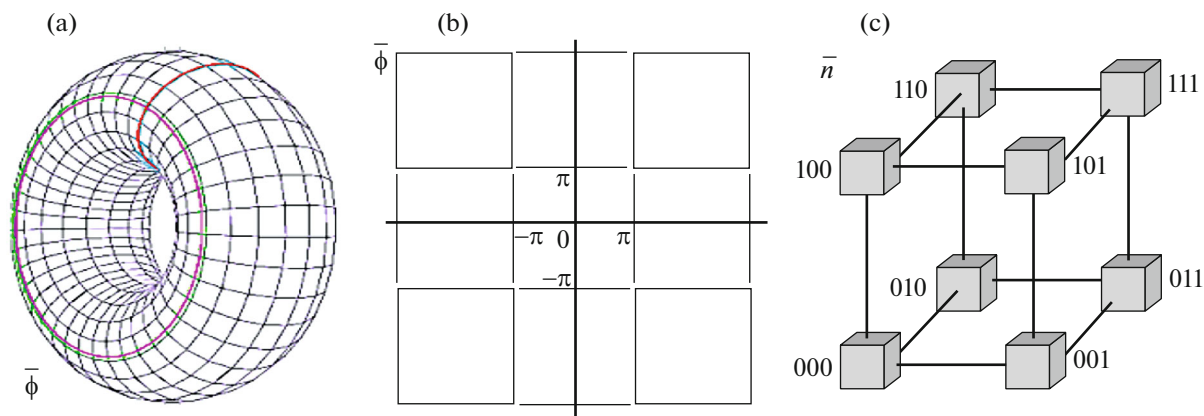
Fig. 2. Schematic representation of: (a) topology of the configuration space of torsion angles (for two dimensions); (b) flattening of a torus in Euclidean space (an example of two-dimensional cells with periodic boundary conditions); and (c) vector space of harmonic numbers (for three dimensions), a single cell is shown.

properties of these functions [27]. A function set on a hypertorus can be expanded into a multidimensional discrete Fourier series [28]:

$$U(\overline{\varphi}) = \sum_{\overline{n}} A(\overline{n}) e^{i\overline{n}\,\overline{\varphi}}, \qquad (25)$$

where we introduced the algebraic vector of the set of torsion angles

$$\overline{\varphi} = (\varphi_1, \varphi_2, ..., \varphi_i, ..., \varphi_N) \qquad (26)$$

and the algebraic vector of the corresponding harmonic numbers of the Fourier series

$$\overline{n} = (n_1, n_2, ..., n_i, ..., n_N); \quad n_k = 0, \pm 1, \pm 2. \qquad (27)$$

The scalar product of vectors is defined in the usual way:

$$\overline{n}\overline{\varphi} = \sum_{k=1}^{N} n_k \varphi_k. \qquad (28)$$

The expansion coefficients (25) are determined by the integral

$$A(\overline{n}) = \frac{1}{(2\pi)^N} \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} U(\varphi) e^{-i\overline{n}\,\overline{\varphi}} d^N \overline{\varphi}. \qquad (29)$$

Note that the familiar interatomic potentials have a singularity at the zero point. However, this singularity lies in the classically forbidden region and does not prevent us from using the Fourier series expansion for the potential energy surface (PES). This circumstance is considered in more detail in [22]. Note also that expansion, for example, in a three-dimensional Fourier series is often used to calculate the contribution of Coulomb interactions when simulating the dynamics of systems with periodic boundary conditions (Ewald sums).

The expansion coefficients in expression (25) contain all the information about the structure of the multidimensional PES, which is associated with the fea-

tures of the chemical structure of the macromolecule, its electronic state, and the properties of its environment. The complex form of the expansion representation was chosen for convenience. Since the potential energy function is real, we have obvious relations for the expansion coefficients (below in the text we show multidimensional algebraic vectors in bold, and in the formulas we keep the notation with an overline):

$$A(\overline{n}) = |A(\overline{n})|e^{i\theta(\overline{n})}; \quad A(-\overline{n}) = A(\overline{n})^* = |A(\overline{n})|e^{-i\theta(\overline{n})}. \qquad (30)$$

The complex amplitude argument $A(\mathbf{n})$ in expressions (30) varies within

$$-\pi < \theta(\overline{n}) \leq \pi. \qquad (31)$$

Since the sum in expression (25) is real, we also conclude that the amplitudes and phases of the expansion coefficients are related by the conditions

$$|A(-\overline{n})| = |A(\overline{n})|; \quad \theta(-\overline{n}) = -\theta(\overline{n}). \qquad (32)$$

The phase relation θ can also be represented as the scalar product of the vector of harmonic numbers $\mathbf{n}$ (27) and some vector $\mathbf{C}(\mathbf{n})$, which is an even function of the vectors of harmonic numbers:

$$\theta(\pm\overline{n}) = \pm\pi h(\overline{n}) - \overline{C}(\overline{n})\overline{n}. \qquad (33)$$

Here, $h(\mathbf{n})$ takes the value 0 if the phase θ lies in the right complex half-plane (the real part of the amplitude in expression (30) is positive) or 1 otherwise.

These almost obvious relations will be very useful for understanding the general properties of the energy surface (25) set on a multidimensional torus of a set of rotation angles about chemical bonds. Since in the sum in expression (25) the terms for the vectors $\mathbf{n}$ and $-\mathbf{n}$ always occur as a pair, we obtain

$$U(\overline{\varphi}) = \sum_{\overline{n}} (-1)^{h(\overline{n})} |A(\overline{n})| \cos[\overline{n}(\overline{\varphi} - \overline{C}(\overline{n}))]. \qquad (34)$$

Note that, in expression (34), we took into account relations (32) and (33) and the fact that, when summing over all possible vectors **n** in expression (34), each term in expansion (25) occurs twice. The expansion coefficients $A(\mathbf{n})$ and the phases $\theta(\mathbf{n})$ terms of the series determine, generally speaking, a very complex structure of the multidimensional potential energy surface with a large number of critical points [29].

It should be noted that such a representation of the PES fundamentally differs from the PES calculation methods that are commonly used in molecular modeling using atom−atom potentials and slightly resembles the use of Ewald sums [30, 31]. Careful calibration of force field parameters for specific types of molecular structures often makes it possible to achieve good results and a certain agreement between experimental and calculated data. However, as already noted [20], the calculation of the PES by summing the contributions of atom-atom potentials has an obvious internal contradiction associated with the accuracy of calculations. For a system of, for example, 1000 atoms, we have about one million terms, each of which has an error of the order of a percent. Therefore, the accuracy of calculating PES (and even more so, PES gradients) is a big question in the general case. At the same time, paradoxically, the molecular dynamics methods for large molecules still work, despite the fact that the accuracy of determining the entire PES and its gradients is poorly controlled. This paradox can be explained by the fact that, according to the general laws for conformational motions in a viscous medium [19] and the results of the first part of the article, only relatively small regions of the configuration space are accessible, and for molecular modeling, PES approximation is required only on very limited and relatively smooth parts of PES [19]. Such an approximation apparently becomes possible due to the calibration of the force field parameters for certain types of systems. (Note that, in many calculations, the thermostat serves as a viscous medium). On the other hand, it is clear that this variant of PES calculation is not suitable for setting the problem in studying general issues related to the formation of the spatial structure and dynamic properties of macromolecules. In this case, we must strictly adhere to an internally consistent scheme for describing PES by using basic principles. The use of the above multidimensional Fourier expansions seems too complicated (and, in general, it is) compared to the use the sum of point atom-atom potentials. On the other hand, the representation of the energy landscape as a sum of such potentials does not make it possible to analyze the general patterns of organization of ultra-multidimensional PES, which are associated with the topology of the configuration space, the symmetry of systems with respect to permutations of identical links, possible chemical evolution of the structures of linear polymers, etc. From the standpoint of the commonly accepted methods of representing PES as a sum of pair potentials, the appearance of macromolecular structures with PES that have a single global minimum and satisfy the folding conditions in a reasonable time seems very strange or a game of improbable chance. Moreover, there have been no ideas about the topography of the PES in that case. Below, using the analytical structure of PES described by expression (25), we discuss the fundamental possibility of realizing relatively smooth energy surfaces with a clearly defined single global minimum during the formation, selection, and evolution of linear polymers.

We note an important feature of PES defined by Eq. (34) in the case of mirror-symmetric structures (i.e., linear polymers without chiral units). Since, in mirror reflection, the signs of the angles change to opposite, for mirror-symmetrical structures we have

$$U_s(\overline{\varphi}) = U_s(-\overline{\varphi}). \tag{35}$$

In this case, the expansion coefficients in expression (25) are real, and in formulas (33) and (34) we must put

$$\overline{C}_s(\overline{n}) = 0 \tag{36}$$

for all harmonic vectors of the multidimensional Fourier expansion (25). In this case, the expansion for PES is reduced to the form

$$U_s(\overline{\varphi}) = \sum_{\overline{n}} (-1)^{h(\overline{n})} |A(\overline{n})| \cos[\overline{n}\overline{\varphi}]. \tag{37}$$

Among the many theoretically possible potential surfaces described by expressions (25) or (34), a class of surfaces for chiral linear polymers that may have a single clear global minimum corresponding to a unique spatial structure can be constructedd. For such surfaces, the phase value $\theta(\mathbf{n})$ in the expansion terms can be selected, for example, as follows [20]:

$$\theta(\pm\overline{n}) = \pm\pi - \overline{n}\overline{\varphi}_m, \tag{38}$$

where $\overline{\varphi}_m$ is the vector of torsion angles for the global minimum position. In this case, relations (32) and (33) for the phases $\theta(\mathbf{n})$ obviously hold. In this case, the potential energy surface can be written as

$$U(\overline{\varphi}) \sim \sum_n |A(\overline{n})| \cos[\overline{n}(\overline{\varphi} - \overline{\varphi}_m)]. \tag{39}$$

A necessary condition for this analytical representation of the PES, as already noted, is the chirality of the polymer molecule. If the molecule was mirror-symmetric, then it should have at least a second symmetric minimum with coordinates at the point $-\overline{\varphi}_m$. We also note that, in the case of mirror reflection, when the signs of all angular variables change and we

proceed to a structure with a different chirality, the value of potential energy (39) does not change due to the fulfillment of condition (32) for the absolute value of the expansion coefficients and the cosine parity.

The analytical structure of the potential surface (39) is optimal, among others, for describing a unique spatial structure in terms of the phenomenon of folding, for example, polypeptide chains into unique spatial protein structures. Below, the surfaces for which there is an analytic representation (39) will be referred to as ideal potential energy surfaces (IPES). Possible causes for the formation and selection of such IPPEs in molecular evolution will be discussed below. For real systems, the sum of expression (39), of course, may contain terms that do not fit into the ideal picture of the structure of the phases of the expansion coefficients (38). However, the general structure of the PES is preserved if the contribution of such terms is not decisive.

Previously we showed that under certain conditions for the expansion coefficients in expression (39) we can obtain a rational explanation for various experimental facts that are observed during the folding process and that, at first glance, are in no way related to each other [20, 21]. An analysis of the general structure of the energy landscapes of linear polymers set by expression (34) shows the conditions under which a quite definite physical pattern leading to such energy landscapes can arise. In terms of the general structure of the energy surface, we could think that the physicochemical stage of the evolution of linear polymers led to the maximum simplification of the dependences $h(\mathbf{n})$ and vectors $\mathbf{C}(\mathbf{n})$ and to the transformation of these functions into constants for the most significant terms of expansion (34):

$$\bar{C}(\bar{n}) = \bar{C}_m; \quad h(\bar{n}) = 1. \quad (40)$$

We will consider the possible causes for the formation of polymeric structures with energy landscapes that could be precursors of PES with a structure of type (39). We consider linear polymers whose structure is completely symmetrical with respect to the permutation of monomer units in the chain. Obviously, we are talking primarily about the cyclic homopolymers of the following type:

$$-L - L - \ldots \ldots - L -. \quad (41)$$

Together with these structures, very long almost homopolymer chains can also be taken into consideration. In this case, we can assume that the system is completely symmetrical with respect to the permutation of identical monomeric units away from the ends of the chain. As far as we know, this type of symmetry has not been considered before in molecular biophysics. Each monomer unit may also contain several internal elements with corresponding conformational degrees of freedom (as, for example, in amino acid res-

idues or in identical sequences of amino acid residues). For a more compact presentation, we will denote the set of k internal torsion angles in the monomer L_i by the algebraic vector

$$\bar{\psi}_i = (\psi_{i1}, \psi_{i2}, \ldots, \psi_{ik}). \quad (42)$$

The algebraic vector of harmonic numbers corresponding to this set of torsion angles will be denoted as follows:

$$\bar{m}_i = (m_{i1}, m_{i2}, \ldots, m_{ik}). \quad (43)$$

In this case, the algebraic vector of torsion angles of the polymer chain is defined as

$$\bar{\phi} = (\bar{\psi}_1, \bar{\psi}_2, \ldots, \bar{\psi}_N), \quad (44)$$

and, accordingly, the vector of harmonic numbers for the Fourier expansion of the potential energy is defined as

$$\bar{n} = (\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_N). \quad (45)$$

In the case of a cyclic homopolymer, the potential energy is symmetrical with respect to permutations of the torsion angles of any monomer units:

$$U(\bar{\psi}_1, \bar{\psi}_2, \ldots, \bar{\psi}_i, \ldots, \bar{\psi}_j, \ldots, \bar{\psi}_N)$$
$$= U(\bar{\psi}_1, \bar{\psi}_2, \ldots, \bar{\psi}_j, \ldots, \bar{\psi}_i, \ldots, \bar{\psi}_N). \quad (46)$$

In this case, the expansion coefficients of the potential energy in a Fourier series are invariant with respect to any permutation of harmonic numbers

$$|A(\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_i, \ldots, \bar{n}_j, \ldots, \bar{n}_N)|$$
$$= |A(\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_j, \ldots, \bar{n}_i, \ldots, \bar{n}_N)|;$$
$$\theta(\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_i, \ldots, \bar{n}_j, \ldots, \bar{n}_N) \quad (47)$$
$$= \theta(\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_j, \ldots, \bar{n}_i, \ldots, \bar{n}_N).$$

Note that, when condition (46) is satisfied, the coordinates of the critical points [29] (local and global minima, maxima, and saddle points of different indices) will correspond to different helical configurations, since the corresponding angular variables at these points take the same numerical values. Formulas (46) and (47) also show that when the components in the vector of harmonic numbers are permutated, the values of the scalar product of the vectors $\mathbf{nC}(\mathbf{n})$ and the values of $h(\mathbf{n})$ in expression (33) do not change.

We divide the entire set of vectors $\mathbf{n}$ into non-intersecting subsets $P_i(\mathbf{n})$, which, for some vector $\mathbf{n}$, unite the type of vectors that differ from each other only by permutation of components. Then we can rewrite expansion (34) for homopolymers in the following form:

$$U(\bar{\phi}) = \sum_i (-1)^{h(n_i)} |A(\bar{n}_i)| \sum_{\bar{n} \in P_i(\bar{n})} \cos[\bar{n}(\bar{\phi} - \bar{C}_i(\bar{n}))]. \quad (48)$$

where $\mathbf{n}_i$ is one of the vectors from the set $P_i(\mathbf{n})$. The elements of the set of vectors $P_i(\mathbf{n})$ over which the internal summation is carried out differ in the permutation of the components of the given vector of harmonics $\mathbf{n}$. If all components of the vector of harmonics $\mathbf{n}$ are different, then the number of such elements is equal to $N!$. If there are matching numbers of harmonics in the vector $\mathbf{n}$, then the number of elements N_i is obtained by dividing $N!$ by the product of the factorials of the number of corresponding identical components k_i of the vector $\mathbf{n}$:

$$N_i = \frac{N!}{k_1!k_2!...}. \qquad (49)$$

The maximum possible number of groups of identical numbers of harmonics is $N/2$ different pairs of identical numbers (if N is even). If N is odd, then it is the integer part of $N/2$ pairs. If the given harmonic number occurs in the vector only once, then $k_i = 1$. If the given harmonic number occurs N times (a vector of the form $(n, n, n,..., n)$), then $k_1 = N$, the rest $k_i = 0$, etc.

The vector $\mathbf{C}_i(\mathbf{n})$ in Eq. (48) is arranged as follows. When the components of the vector of the numbers of harmonics $\mathbf{n}$ are permutated, the components of the vector $\mathbf{C}_i(\mathbf{n})$ are also rearranged so that the scalar product of these vectors does not change. For a cyclic homopolymer built from nonchiral monomers, the vector $\mathbf{C}_i(\mathbf{n})$ is just equal to 0. Note also that the inner sum in Eq. (48) is symmetric with respect to the permutation of the components of the torsion angle vector (44).

In a general case, the inner sum in Eq. (48) contains a very large number of oscillating terms with the same amplitude whose phases are not correlated. This sum does not allow clear extrema to be distinguished, and a linear homopolymer built even from chiral monomers can form a unique spatial structure only under specific conditions. This condition can be clearly seen from Eq. (48). To form a clear global minimum, it is necessary that the components of the vector $\mathbf{C}_i(\mathbf{n})$ be the same and do not depend on the permutations of components in the vector $\mathbf{n}$. In this case, a phase correlation in the inner sum of Eq. (48) occurs, and a sharp increase in the extremum (proportional, maximum, by a factor of $N_i!$ times) in the configuration determined by the vectors $\mathbf{C}_i(\mathbf{n})$ with the same set of distinct components is observed.

Since these vectors in the space of angles have the same (for each type of torsion angle) components, the corresponding conformation is a helix. Thus, if the chemical structure of the homopolymer chain is such that the potential energy during the formation of the helix decreases, then the internal sum in Eq. (48) gives a clear minimum in PES. The appearance of this helix can be very diverse, especially if the monomeric unit contains many internal conformational degrees of freedom. In addition, if the polymer chain gains some energy when several types of helices are formed, denoted below as α, β, etc. (coincidence of designations with the known types of helices in polypeptides in this case does not carry any semantic load), then we can represent the sum (48) in the form

$$U(\overline{\varphi}) \sim - \sum_{i=\alpha,\beta,...} |A(\overline{n}_i)| \sum_{n\in P_i(n)} \cos[\overline{n}(\overline{\varphi} - \overline{C}_i)]$$
$$+ \sum_{i\neq\alpha,\beta,...} (-1)^{h(n_i)} |A(\overline{n}_i)| \sum_{n\in P_i(n)} \cos[\overline{n}(\overline{\varphi} - \overline{C}_i(\overline{n}))], \qquad (50)$$
$$\overline{C}_i = (\overline{C}_{im}, \overline{C}_{im},..., \overline{C}_{im}).$$

Recall that the vector notation for the components $\overline{C}_{im}$ is related to the fact that it reflects the set of values of internal torsion angles in monomers. The first group of terms is the contributions to the PES from helical structures, each of which has a well-defined phase in the conformation space and a negative sign. The second group of terms does not have a clearly defined phase and sign and is seen rather as noise. This leads us to the conclusion that, due to the symmetry with respect to the permutation of identical monomeric units, homopolymer chains create the background (at a certain chemical structure) for the formation of unique helical spatial structures. This effect might be the first stage in the evolution of energy landscapes, which may be the cause for the abundance of helical structures (as well as ring structures in the simplest organisms) in the spatial organization of biopolymers. We do not associate the discussed possible types of protoforms of linear polymers with proteins or nucleic acids in living systems. Modification and adjustment of the chemical forms of monomers apparently occurred over the course of long-term physicochemical evolution.

Let us further consider what we can obtain from the above symmetry considerations for understanding the protein folding processes, which are still largely mysterious.

We proceed from the generally accepted concept that, for a macromolecule that forms a unique spatial structure, PES has a global minimum and the folding process is designed so that a representative point has the ability to reach a global energy minimum. We also consider that the general analytical form of the multidimensional PES is close to the structure that was conditionally called ideal (IPES) and is described by Eq. (39). Above, we expressed considerations in favor of the selection of such forms of PES in physicochemical evolution, which were based on symmetry considerations with respect to the permutation of identical (or almost identical) monomeric units. On the other hand, it would be very strange if an extensive class of linear biopolymers with a unique spatial structure (for example, proteins that are synthesized according to common physicochemical mechanisms) were represented by objects in which the global PES minimum is formed randomly for each object. Previously, we showed [20, 21] that, using the IPES structure

described by Eq. (39), we can obtain results that allow interpretation of the most diverse effects for protein folding.

Now we consider the dependence of the series expansion coefficients in Eq. (39) on the harmonic number vectors. Today, there are general considerations only as to how these expansion coefficients should be arranged. As we noted above, in the expansions for PES (Eq. (48)) or in the regular part of the expansion (50), due to the symmetry with respect to the permutation of identical monomer units, the coefficients of the series must be functions of some invariants of the harmonic number vectors (i.e., functions that do not change when the components of vectors **n** for identical chain links are permutated). The simplest invariants are the length of the vector or, for example, the scalar product of the vector **n** and a vector proportional to the unit vector. The last case in a more general variant was discussed by us earlier in [20, 21]. Below, we consider a model of the Gaussian dependence of the series expansion coefficients on the length of the harmonic number vector, which gives results qualitatively similar to those obtained in [20, 21] but allows more compact equations from which interesting consequences can be seen.

In the simplest case, when there is only one torsion angle in the monomer, we will proceed from the following Gaussian form for the expansion coefficients:

$$|A(\overline{n})| = a_0 \frac{1}{2\lambda\sqrt{\pi}} \left\{ \exp\left[-\frac{1}{\lambda^2}(\overline{n} - \overline{a})^2 \right] + \exp\left[-\frac{1}{\lambda^2}(\overline{n} + \overline{a})^2 \right] \right\} \exp\left[-\sum_i \varepsilon_i n_i \right]. \quad (51)$$

The expansion coefficients in Eq. (51) will be symmetrical with respect to permutations of the components of the harmonic number vector if the multidimensional algebraic vector **a** is equal to zero or proportional to the unit vector. Below we will see that all equations also work for some violation of ideal symmetry in the case of an arbitrary vector **a**. If there are k torsion angles in monomeric links, then the formula becomes somewhat more complicated in accordance with the notation in Eqs. (42)–(45):

$$|A(\overline{n})| = \frac{a_0}{2^k \pi^{k/2}} \left\{ \sum_{j=1}^{k} \frac{1}{\lambda_j} \exp\left[-\frac{1}{\lambda_j^2}(\overline{m}_j - \overline{a}_j)^2 \right] + \sum_{j=1}^{k} \frac{1}{\lambda_j} \exp\left[-\frac{1}{\lambda_j^2}(\overline{m}_j + \overline{a}_j)^2 \right] \right\} \exp\left[-\varepsilon \sum_i |n_i| \right]. \quad (52)$$

Let us explain the meaning and capabilities of these equations. First of all, the last factor reflects the fact that, for a physically reasonable function, the coefficients of the Fourier series decrease approximately exponentially with increasing numbers of harmonics [22, 28]. The decay rate of the coefficients with increasing harmonic number is determined by the ε parameters. Below, in the final equations for the

PES, we proceed to the limit $\varepsilon \to 0$, and we do not care about the differences in this parameter for different degrees of freedom.

The sum of two Gaussian exponents and the sign of vector **a** in the argument of the second exponent are introduced to ensure the symmetry of the expansion coefficients when the sign of the harmonic number vector changes in accordance with the general requirements for Eqs. (30) and (32). Note that series (39) is summed over all values of harmonic numbers, both positive and negative. Therefore, the addition of the second Gaussian exponent in Eqs. (51) and (52) is decorative. The λ parameters determine the scale of acceptable deviations in values from 0 in Gaussian exponents. When λ tends to 0, the Gaussian exponents turn into delta functions. Despite the relative simplicity of Eqs. (51) and (52), they make it possible to generate very diverse types of IPES by varying the values of the introduced parameters. Importantly, such surfaces can be studied by analytical methods. Previously, the capabilities of dependences similar to those described by Eqs. (51) and (52) for characterizing the topography of the PES and the topography of the free energy surface were studied in detail [20–22]. The variant proposed below is more convenient for discussing the general effects of symmetry on PES topography with respect to rearrangement of monomer units in the polymer chain.

The convenience of equations such as (51) for obtaining analytical results is based on the use of the integral representation

$$\exp[-(n_i \mp a_i)^2/\lambda^2] = \frac{\lambda}{2\sqrt{\pi}} \times \int_{-\infty}^{\infty} \exp\left[-\frac{\lambda^2}{4} t^2 \pm i(n_i \mp a_i)t \right] dt, \quad (53)$$

and using the sum of the series

$$F_0(\phi, \varepsilon) = \sum_{n=-\infty}^{\infty} e^{-\varepsilon|n| \pm in\phi} = \frac{\sinh \varepsilon}{\cosh \varepsilon - \cos \phi} \sim \frac{2\varepsilon}{\varepsilon^2 + 4\sin^2(\phi/2)}; \quad \varepsilon \ll 1. \quad (54)$$

If we substitute the coefficients from Eq. (51) into Eq. (39), select the representation for the cosines in Eq. (39) as a sum of complex exponents, represent the Gaussian exponents in Eq. (51) as a product of factors from Eq. (53), take into account that the sum of all possible products is equal to the product of the sums, and use Eq. (54), we obtain the following expression for the PES:

$$U(\overline{\phi}) = -\frac{a_0}{2} \frac{\lambda^N}{2^N \pi^{N/2}} \prod_{i=1}^{N} \int_{-\infty}^{\infty} \exp\left[-\frac{\lambda^2}{4} t_i^2 - ia_i t_i \right] \times F_0(t_i + \phi_i + \phi_{mi}, \varepsilon) dt_i + CC. \quad (55)$$

CC in Eq. (55) denotes the complex conjugate term. For $\varepsilon \ll 1$, the main contribution to the integrals comes from the vicinity of zero for the argument of function F_0, and we obtain

$$U(\bar{\phi}) \sim a_0 \lambda^N \pi^{N/2} e^{-\frac{\lambda^2}{4}\sum_i (\phi_i - \phi_{im})^2}$$
$$\times \cos \sum_i a_i(\phi_i - \phi_{im}); \quad \lambda\varepsilon \ll 1. \tag{56}$$

The structure of expression (56) shows that it is advisable to introduce two generalized variables

$$R^2 = \sum_i (\phi_i - \phi_{im})^2 = (\bar{\phi} - \bar{\phi}_m)^2;$$

$$Z = \frac{1}{\|\bar{a}\|}\sum_i a_i(\phi_i - \phi_m) \tag{57}$$

$$= \frac{\bar{a}}{\|\bar{a}\|}(\bar{\phi} - \bar{\phi}_m), \quad \|\bar{a}\|^2 = \sum_i a_i^2.$$

Geometrically, the first variable defines hyperspheres of radius R in the space of torsion angles described around the global minimum point. The second variable determines the hyperplanes perpendicular to the vector **a** and located at a distance Z (with accuracy of up to a sign) from the global minimum point.

As a result, the PES described by Eq. (56) can be rewritten in a simple form:

$$U(\bar{\phi}) = U(R, Z) = -U_0 e^{-\frac{\lambda^2}{4}R^2} \cos\|\bar{a}\|Z; \quad R \geq |Z|. \tag{58}$$

Compliance with the nonstrict inequality $R \geq Z$ in Eq. (58) is very important, since otherwise there is no such set of angular variables that simultaneously satisfy conditions (57). This is clearly seen from Fig. 3.

If we do the same with Eq. (52), we obtain a slightly more complex formula:

$$U(\bar{\phi}) = U(R, Z) = -U_0 e^{-\frac{\lambda^2}{4}R^2} \cos\|\bar{a}_\lambda\|Z; \quad R \geq |Z|. \tag{59}$$

The structure of Eq. (58) was retained, but the parameters included in it took the following form:

$$R^2 = \sum_{i=1}^{N}\sum_{j=1}^{k}(\xi_{ij} - \xi_{ijm})^2;$$

$$Z = \frac{1}{\|\bar{a}_\lambda\|}\sum_{i=1}^{N}\sum_{j=1}^{k}\frac{a_{ij}}{\lambda_j}(\xi_{ij} - \xi_{ijm}); \tag{60}$$

$$\xi_{ij} = \lambda_j \psi_{ij}.$$

Here the index i numbers the monomers in the chain, and the index j numbers the types of torsion angles in the monomer. It can be seen that the general principles of the arrangement of the PES topography are retained. Using scaling transformation of the angular variables, we can proceed to the familiar situation of section of a hypersphere of dimension $N*k$ with a
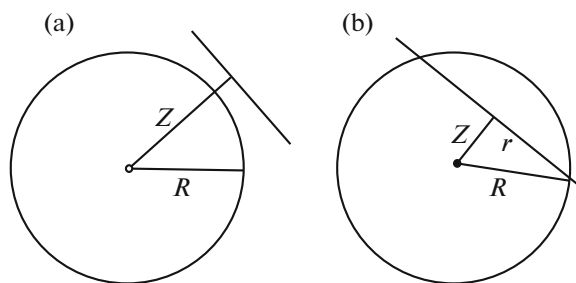

Fig. 3. Explanation for the fulfillment of the condition $R > Z$: (a) the radius of the sphere is less than the distance to the plane and there are no points of intersection of the angular variables that satisfy Eq. (58); (b) the radius of the sphere is greater than the distance to the plane and there is a set of intersection points that satisfy Eq. (58), which is a sphere with dimensionality smaller by 1 and radius $r = (R^2 - Z^2)^{1/2}$. If $R = Z$, then there is a point of contact.

hyperplane located at a distance Z from the center (the point of the global minimum of the PES), which is perpendicular to a vector with components (a_{ij}/λ_j), and obtain similar restrictions on the R values. The designations are clear from the context. We can also trace the main characteristics of the PES topography on a simpler variant of the surface given by Eq. (58).

Typical examples of PES topography in the case of two angular variables are shown in Fig. 4. If the vector **a** is equal to 0, then we just have a Gaussian funnel. For a finite length of the vector **a**, the PES relief is a set of funnels. There is the deepest central funnel, which is surrounded by satellite funnels whose depth decreases with distance from the central funnel. All funnels are separated from each other by potential barriers.

The considered formulas have the following useful property: even if there are more than two angular variables, we can still study the PES topography in terms of the generalized variables R and Z.

An example of the PES topography in terms of the R and Z variables is shown in Fig. 5. We can see that, at a finite length of the vector **a**, the space of configurations of the polymer chain is cut into regions with relatively low conformational energy, which are separated from each other by potential barriers (Fig. 6).

As in the case described in [20], the PES topography is a system of energy funnels (Fig. 7). The generalized variables R and Z have a central and deepest funnel that leads to the global PES minimum. There are also many satellite funnels whose depth decreases with distance from the central funnel. Satellite funnels in generalized variables, as well as in angular variables, are separated from the central funnel and from each other by potential barriers.

Lets further discuss the role and value of the vector **a** in terms of the considered Gaussian approximation. If the polymer under consideration consists of absolutely identical monomeric units and the surface is
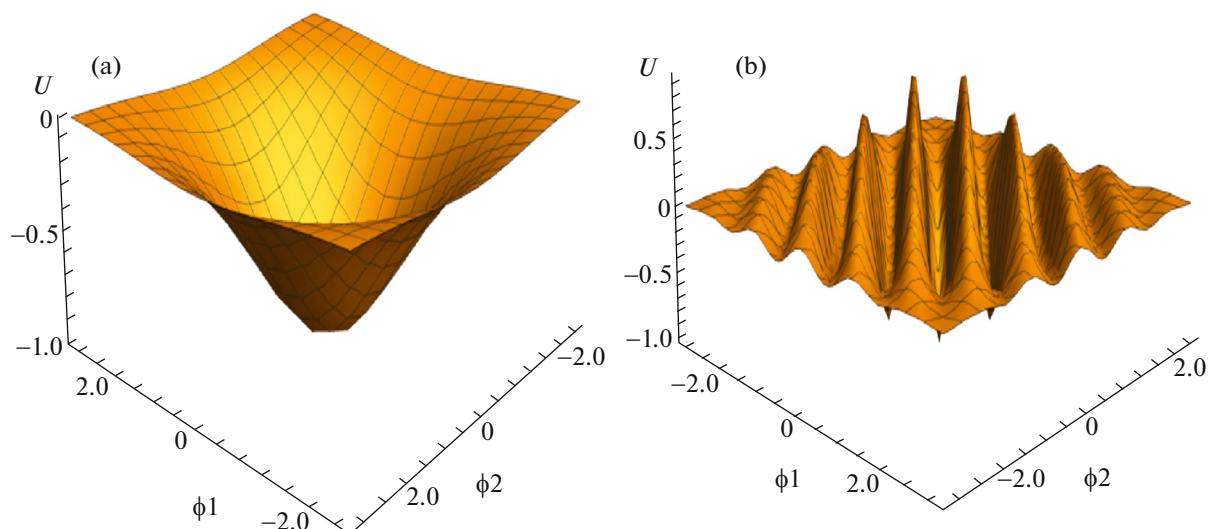
Fig. 4. Characteristic PES topographies in the Gaussian approximation for two angular variables $U(\bar{\phi}) = -\exp\left[-\frac{1}{4}(\phi_1^2 + \phi_2^2)\right] \times \cos[\|a\|(\phi_1 + \phi_2)]$: (a) $- \|a\| = 0$, (b) $- \|a\| = 5$.

symmetrical with respect to the permutation of torsion angles of the same type, then, for symmetry causes, this vector should either be equal to zero or be proportional to the unit vector. In the first case, we have a PES with a simple topography, as in Fig. 4a. This topography of the PES fully complies with the principle of minimum frustration and is ideal for very fast folding into a unique spatial structure (in this case, helical) that is achievable from any initial chain configuration. It should be noted that examples of very fast folding of polypeptides over nanoseconds are known [12]. If the vector **a** is nonzero, then the topography of the set of nested energy funnels is formed (Fig. 4b). In this case, certain features of the process of

polymer folding are manifested. Inside the central funnel (as well as inside the satellite funnels), the topography of the energy surface satisfies the principle of minimum frustration. The coordinates and energy value in the minimum of the satellite funnel are determined according to Eq. (58) by the following relationship:

$$\|\bar{a}\|Z = \sum_i a_i(\phi_i - \phi_{im}) = 2\pi k; \quad k = \pm 1, \pm 2... \quad (61)$$
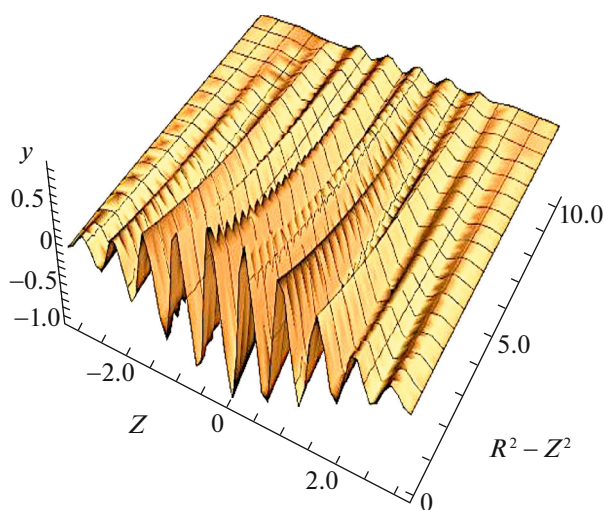

Fig. 5. PES topography according to Eq. (58) in generalized variables Z and $R^2 - Z^2$. Model PES of the form $U(R, Z) = -\exp(-0.25R^2)\cos(10Z)$.
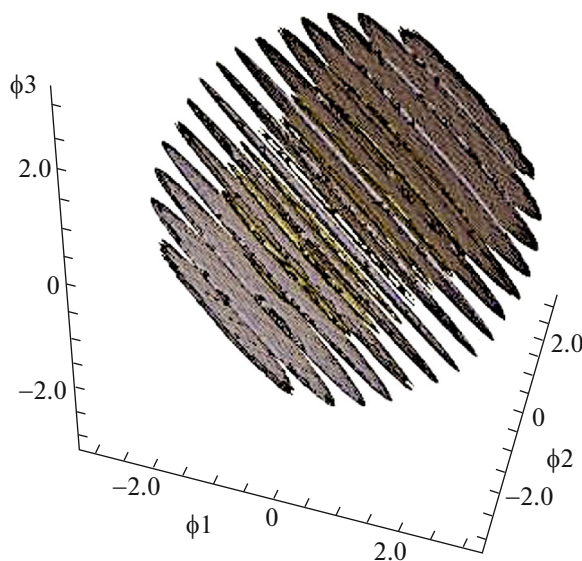

Fig. 6. Spatial section of a three-dimensional PES. An example of stratification of the configuration space into energy-allowed (in the vicinity of local minima) and energy-forbidden (not shaded in the figure) layers for a PES of a model type (according to Fig. 5).

The value of the R coordinate at the minimum point is determined according to (58) by the value $R = Z$ (or $R = -Z$). Accordingly, the energy values in the minima of satellite funnels are determined by the equation

$$U_k = -U_0 \exp(-\pi^2\lambda^2 k^2/\|\overline{a}\|^2)e^{\frac{-\lambda^2}{4}R^2}; \quad R \geq 2\pi k. \quad (62)$$

The shifts of the coordinates of the positions of the energy minima in satellite funnels relative to the position of the global minimum are

$$\phi_{ik} - \phi_{im} = 2\pi k a_i/\|\overline{a}\|^2. \quad (63)$$

Note that the number of satellite funnels is limited by obvious conditions:

$$|\phi_{ik} - \phi_{im}| < \pi; \quad |k| < \|\overline{a}\|^2/2|a_i|. \quad (64)$$

The energy values for the height of potential barriers between the funnels k and $k - 1$ (we consider only positive $k \geq 1$, for negative k the situation is symmetrical) are determined by the expressions

$$B_k(R) = U_0 \exp[-\pi^2\lambda^2(2k-1)^2/4\|\overline{a}\|^2]e^{\frac{-\lambda^2}{4}R^2}; \quad (65)$$
$$R \geq (2k-1)\pi.$$

Subtracting from the barrier height (65) the value of the energy in the minimum of the satellite funnel (62), we see that the possibility of a representative point passing between adjacent funnels increases rapidly with an increase in the funnel number and the value of the generalized coordinate R. An increase in the length of the vector **a** contributes to a sharper PES topography with more pronounced barriers and deeper minima. An increase in the λ parameter, conversely, smoothes the PES topography without changing the central funnel depth.

Thus, within the Gaussian model considered, there is a vast range of parameter values for which the transition of a representative point between energy funnels is very difficult. This PES structure includes a mechanism of connection between the initial chain conformation, which arises in the course of synthesis, and the final spatial structure obtained in the process of folding for a reasonable time. From the standpoint of the selection of linear polymers that form strictly defined spatial structures, the advantage at the first stage should be given to the macromolecules whose PES is arranged as symmetrically and smoothly as possible. In the framework of the considered Gaussian model, these are structures with PES for which the vector **a** is equal to zero. If the monomeric units of such structures were further modified (e.g., in the region of side groups), then this ideal symmetry for the expansion coefficients and PES is violated. Within the
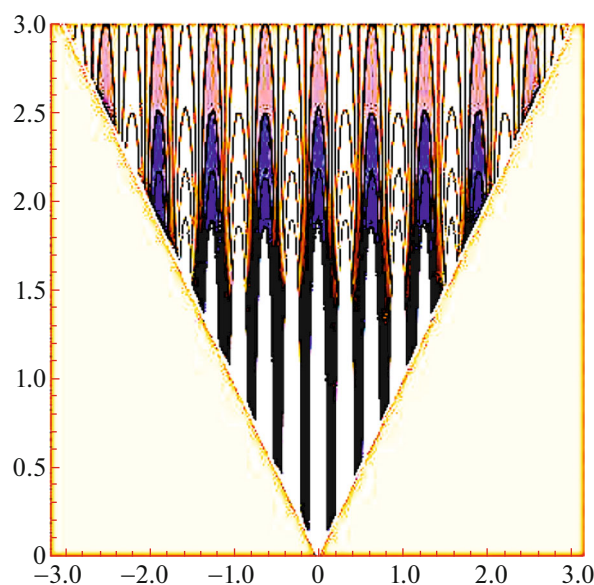

Fig. 7. Energy funnels for PES (according to Fig. 5) in generalized variables R and Z. The Z axis is horizontal, $R^2 > Z^2$. The deepest funnel is the central funnel, on the right and on the left are satellite funnels of decreasing depth. Funnels are separated from each other by barriers whose heights decrease with distance from the global minimum point.

considered model, this is interpreted as the appearance of a nonzero vector **a** and complication of the PES topography. Observing the general symmetry conditions for the expansion coefficients (32), in the Gaussian model we describe these changes as a shift in the distribution center of the expansion coefficients in the space of harmonic numbers by the vectors **a** and −**a** (Fig. 8). This is the minimum step in the complication of the PES topography, which is very similar to removal of degeneracy of a multiply degenerate state by including an additional perturbation. In this case, as we saw above, the PES topography has the form of a system of funnels separated by energy barriers. In this case, the uniqueness of the result of polymer folding for a reasonable time requires certain conditions for the starting configurations of the chain. To successfully collapse and move a representative point to a global minimum, these spatial configurations must belong to a specific, though very broad, set of initial states that are in the area of attraction of the central energy funnel. From the standpoint of the ideas developed, in modern living systems this can apparently be realized in the variant of co-translational folding, when the starting set of configurations of the polypeptide chain is determined by the mechanisms of biosynthesis. Apparently, a role in the formation of such starting configurations is also played by chaperones, which fix certain chain configurations [32] and thereby direct the folding trajectory in the required direction.
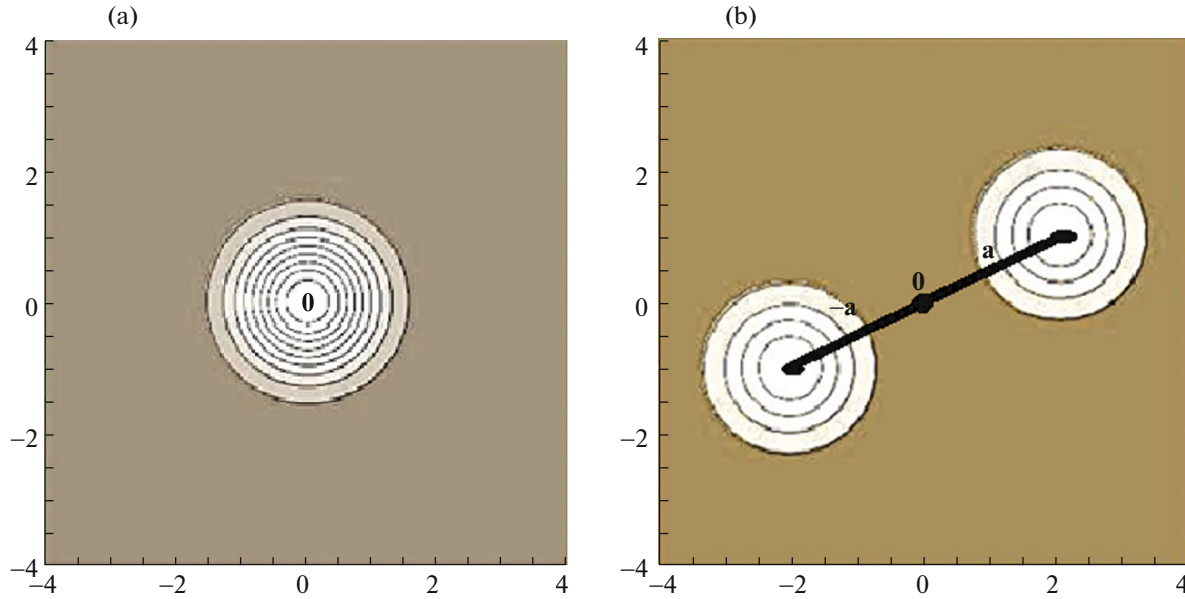
(a)

(b)

Fig. 8. Spread of contributions from regions of the space of harmonic numbers in the case of ideal symmetry **a** = 0 of the polymer chain with respect to permutation of monomer units (a) and weak symmetry (b) in the Gaussian approximation.

FREE ENERGY SURFACE OF A LINEAR POLYMER CHAIN

Now we proceed to the consideration of the free energy surface (FES) of a linear polymer chain, which forms a unique spatial structure. In real experiments on folding at finite temperatures, the observed effects are determined precisely by the free energy surface of macromolecules. Within Gaussian models developed for the PES, there is a just simple and logical transition from the PES to the free energy surface (FES) of the chain in terms of the generalized coordinates R and Z. It is known that the free energy of the system in the state of equilibrium under given conditions (R and Z values) is defined through the sum over the states [33]:

$$G(R,Z) = -k_{\mathrm{B}}T \ln W(R,Z)$$
$$W(R,Z) = \int_{\{\phi\}} \exp[-U(\bar{\phi})/k_{\mathrm{B}}T]\delta(R - \|\bar{\phi} - \phi_m\|) \quad (66)$$
$$\times \delta(Z - \bar{a}(\|\bar{\phi} - \phi_m\|)/\|\bar{a}\|)\theta(R - |Z|)d^N\bar{\phi}.$$

The delta functions in the integral (66) cut out the regions of the N-dimensional configuration space of torsion angles, which are defined by equations (57), and the Heaviside theta function cuts out the values of the variables R and Z in accordance with condition (58). Despite the complex form of integral (66), its calculation within the Gaussian model for PES is not difficult. The delta functions and the Heaviside function, for a given potential energy, cut out the volume of a hypersphere with radius $r = (R^2 - Z^2)^{1/2}$ from the phase space (see Fig. 3), and for the sum over states, respectively, for the free energy we obtain

$$W(R,Z) \sim (R^2 - Z^2)^{(N-1)/2}e^{-U(R,Z)/k_{\mathrm{B}}T} = e^{-G(R,Z)/k_{\mathrm{B}}T};$$
$$R \geq |Z|. \quad (67)$$

respectively, for the free energy we obtain

$$G(R,Z) = U(R,Z)$$
$$- k_{\mathrm{B}}T(N-1)\ln\sqrt{R^2 - Z^2}; \quad R \geq Z. \quad (68)$$

The entropy of a state with fixed R and Z is proportional to the logarithm of the phase volume of these states:

$$S(R,Z) = k_{\mathrm{B}}(N-1)\ln\sqrt{R^2 - Z^2}; \quad R \geq |Z|. \quad (69)$$

We omitted the numerical factor in Eq. (67), since after taking the logarithm, it affects only the origin of the free energy in Eq. (68). There is a problem in Eq. (68) at the point $R^2 = Z^2$, where the free energy becomes minus infinity, since the phase volume of the point of contact between the hypersphere R and the hyperplane Z is equal to zero. In this case, we need to take into account the quantum corrections due to zero oscillations in torsion angles with a finite amplitude. In this case, $(R^2 - Z^2)$ cannot be less than some finite value proportional to the square of the amplitude of zero-point vibrations in torsion angles. Let us estimate these values. We assume that the amplitudes of zero oscillations are the same for all angles; according to Eq. (57), we have:

$$\langle R^2 \rangle_0 = \langle \|\bar{\phi} - \bar{\phi}_m\| \rangle_0 = N\langle \phi^2 \rangle_0$$
$$\langle Z^2 \rangle_0 = \langle \|\bar{\phi} - \bar{\phi}_m\|^2 \cos^2\alpha[\bar{a},(\bar{\phi} - \bar{\phi}_m)] \rangle_0 = \frac{1}{2}N\langle \phi^2 \rangle_0. \quad (70)$$

The angle α in Eqs. (70) is the angle between the vectors; when averaging, the squared cosine is averaged over all orientations. Thus, we obtain

$$\langle R^2 - Z^2 \rangle_0 \simeq \frac{1}{2} N \langle \phi^2 \rangle_0. \tag{71}$$

Now we estimate the squared amplitude of zero-point oscillations over torsion angles following the commonly used procedure [26]. Assuming that the potential energy upon rotation near the minimum takes the form

$$U(\phi) \sim \frac{1}{2} u (1 - \cos 3\phi) \sim \frac{9}{4} u \phi^2, \tag{72}$$

we obtain for the squared amplitude of zero oscillations:

$$\langle \phi^2 \rangle_0 \simeq \frac{\hbar}{3l\sqrt{Mu}}, \tag{73}$$

where M is the mass of the side group that rotates during zero torsion oscillations and l is the length of the bond between the node and this group. Assuming that the mass M is approximately 20 a.m.u., l is approximately 1.5 Å, and u is approximately 3 kcal/mol, we obtain a value of approximately 0.01 square radians for the square of the amplitude of zero-point oscillations.

To eliminate the divergence in the expression for entropy (69), we add to the value of the hypersphere radius under the sign of the logarithm the contribution of the amplitudes of zero oscillations in torsion angles, which is determined from Eq. (71). As a result, we obtain:

$$G(R, Z) = -U_0 \left\{ e^{\frac{-\lambda^2}{4} R^2} \cos \|\bar{a}\| Z \right.$$
$$\left. + \frac{T}{T_0} \ln \left[1 + \frac{\sqrt{R^2 - Z^2}}{\sqrt{\langle R^2 - Z^2 \rangle_0}} \right] \right\}; \quad R \geq Z, \tag{74}$$

where we introduced the characteristic temperature parameter

$$T_0 = U_0 / k_B N, \tag{75}$$

which is proportional to the specific decrease in potential energy during folding per one degree of freedom. An example of a free energy surface is shown in Fig. 9.

From Eq. (74), taking into account Eq. (73), one can estimate the temperature threshold at which the free energy of the folded state will be lower than that of the fully denatured state. Note that the limit value of the parameter R^2 is about $N\pi^2$, and the limit value of the quantity under the logarithm sign does not depend on $N(!)$ and is about 45. In other words, within the considered Gaussian model, for the given parameter values, the folded state becomes thermodynamically favorable at $T < 0.26 T_0$ (Fig. 10). Note that this esti-
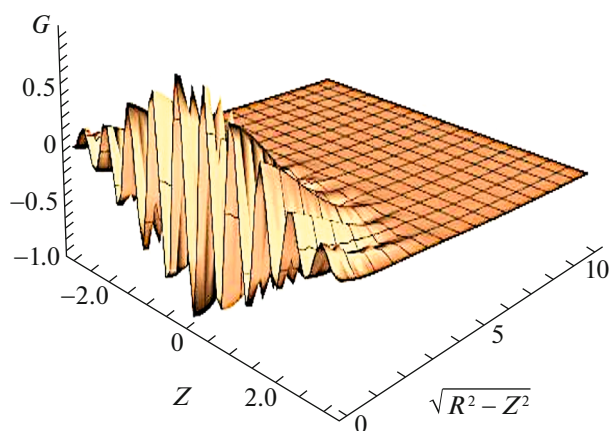

Fig. 9. Topography of the free energy surface (74) in the space of generalized variables $G = -\exp(-0.25R^2)\cos 10Z - 0.15\ln(1 + \sqrt{R^2 - Z^2})$.

mate at large N is almost independent of the length of the chain and the parameter λ.

Assuming, for example, that the temperature threshold for protein denaturation is ~60°C (~333 K), we estimate the parameter T_0 at ~1280 K. Thus, this model predicts, from the denaturation temperature, the gain in the potential energy of the macromolecule per one degree of freedom (in this particular case, it is ~2.56 kcal/mol). We obtain that the energy of nonvalent interactions of two monomer units during folding in this case is ~5 kcal/mol, or close to the hydrogen bond energy in an aqueous medium. Naturally, these are approximate values that show a reasonable order of magnitude.

Figure 10 also shows a characteristic volcano-type free energy profile, which is observed in experiments on protein folding in solutions [9, 10]. The region of the energy funnel that leads to the folded state is separated from the denatured state by a barrier with an entropic nature. In principle, when comparing the experimental value of the entropy barrier height and the dependences in Fig. 10, the range of values of the generalized coordinate R for the state of the denatured protein in solution can be estimated.

POSSIBILITY AND CAUSES OF DIRECTED PHYSICOCHEMICAL MOLECULAR EVOLUTION TOWARDS THE FORMATION OF MACROMOLECULES WITH A UNIQUE SPATIAL STRUCTURE

Further, we note an important circumstance that combines all the above considerations on the topography of PES and FES and the rules for the motion of a representative point along the multidimensional potential energy surface of a macromolecule in a viscous medium, presented in the first part of this work
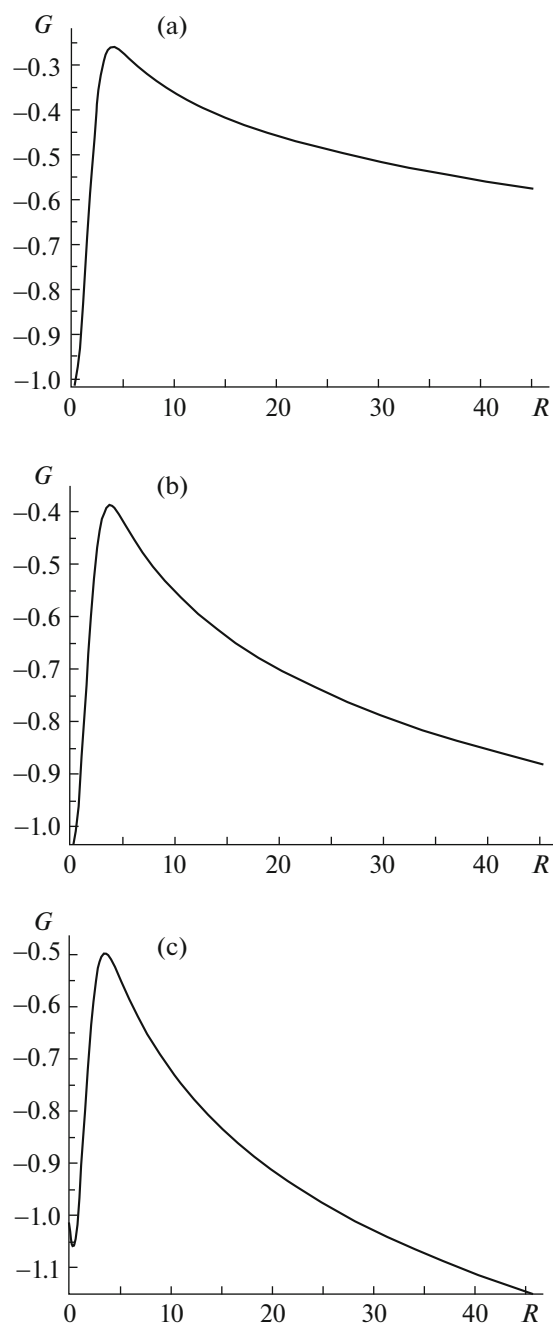
Fig. 10. Free energy profile (according to Fig. 9) in the central funnel ($Z = 0$): (a) $T/T_0 = 0.15$, (b) $T/T_0 = 0.23$, (c) $T/T_0 = 0.3$. (a, b) The folded state is thermodynamically favorable, (c) the folded state is thermodynamically unfavorable.

and additionally in [19]. All the principles outlined above can apparently work in combination and be aimed at solving the problem of molecular evolution—the selection of molecular structures that form unique spatial structures in a specific range of environmental parameters and the chemical composition of polymers.

As noted above, the successful result of folding or getting a representative point in the global minimum of free (or potential) energy in a non-trivial situation, when the vector **a** is not equal to zero, is largely determined by the starting position of the representative point, which is very desirable to be in the region of attraction of the central energy funnel. We see (for example, in Fig. 9) that the FES, as well as the PES, for denatured states far from the bottom of the energy funnels (large R values) is rather smooth, and landing into the central funnel (with the correct selection of the starting configuration in the region of attraction of this funnel) will be possible if the parameter Z does not change too significantly when the representative point moves. Consequently, an almost zero value of the time derivative of the generalized variable Z (especially in the region of large values of the generalized variable R, where the initial configuration of the polymer is formed during the synthesis, and it becomes possible to select the "correct" starting configuration) is a very useful condition for the correct folding process. This means that, in the region of large R and in the zone of attraction of the central funnel, the condition for the start of proper folding is as follows:

$$\frac{dZ}{dt} = \sum_i a_i \dot{\phi}_i \sim 0. \tag{76}$$

Recall that the point above the torsion angle means differentiation with respect to time. Thus, a variant of the folding process when a certain linear combination of angular velocities of rotation around torsion angles is close to zero is very favorable for correct folding. We encountered a similar situation in Eq. (24.1) when discussing the correlation of conformational motions of a polymer chain in a viscous medium:

$$\sum_k \dot{\phi}_k l_k = 0; \quad l_k = (\mathbf{b}_k \mathbf{l})/b_k.$$

Recall that **l** is a vector connecting the beginning and end of the chain, and $\mathbf{b}_k$ is the vector of bonds between the nodes of the chain. For example, if all components of the **a** vector have the same sign (as in polymers that are symmetrical with respect to the permutation of monomeric units), then a variant of an elongated starting conformation, which leads to sign-alternating directions of turns about dihedral angles in accordance with Eq. (24.1) will also be favorable for reaching the representative point to the global minimum. We observed this variant of folding dynamics for a polypeptide chain in a viscous medium [17]. Thus when folding starts from the unfolded chain configuration, polymers that fold in the central funnel into helical structures will gain an advantage.

Equations (76) and (24.1) show the possibility of existence of the effect of a positive correlation between the region of starting chain configurations (arising during synthesis) and the energy surface topography (in the Gaussian model considered, this is defined by

the vector $\mathbf{a}_i$ components), which is determined by the chemical composition and sequence of monomers in the chain. Apparently, favorable realization of this correlation under certain conditions may lead to a tendency for the formation of polymer molecules with unique spatial structures.

CONCLUSIONS

We considered the effects of two types of symmetry that have a nontrivial effect on the dynamics of linear polymers in a viscous medium and on the topography of their energy landscapes.

In the approximation of a homogeneous viscous medium, the dynamics of a macromolecule is invariant with respect to translations and rotations of the object as a whole. This leads to the conservation laws (5) and (13). A less obvious but useful result can be obtained as a result of application of these formulas for a polymer chain with rigid bonds and valent angles. The sum of the vectors of angular velocities of turns around the bonds (or rotations around torsion angles) becomes equal to zero. In principle, this result is not unexpected, since it follows from the equations of motion in a highly viscous medium that the total moment of rotation of the macromolecule must be equal to zero. Equation (24) has additional interpretations if it is multiplied by a specially constructed vector, e.g., a vector connecting the start and end nodes of a linear chain (24.1). The obtained relations clearly show that, at the initial stage of folding of an unfolded linear chain, clear effects of the correlation of turns about dihedral angles with a rather regular change in the sign of angular velocities should be observed. As a result, there is a tendency to form helical structures if such structures are allowed for energy causes. We have previously observed effects of this type when simulating the folding of polypeptide [17, 18] and model polymer [16, 34, 35] structures in a viscous medium using molecular dynamics methods (in certain cases, the role of viscosity can be played, for example, by thermostats [37]). Note that taking into account thermal fluctuations and changing the equations of mechanics for particles in a viscous medium on the Langevin equations [19] makes the chain folding dynamics less deterministic, and the relations discussed are satisfied only on average over the ensemble. However, far from the equilibrium position, in areas with large potential energy gradients, the effect described above may well implement the role of guiding the trajectory of a representative point, for example, towards the formation of helical structures. Let us mention one more important circumstance associated with the rules of motion of a representative point over ultrahigh-dimensional energy surfaces. The restrictions that are imposed on conformational motions in a viscous medium by relations such as (24) and (24.1), together with the restrictions on the distribution of energy dissipation rates over chain nodes, which were described earlier in [19],

make most of the energy surface inaccessible to a representative point, even in the absence of energy prohibitions. This appears to greatly simplify the problem of calibrating the force field parameters, which are used in the calculations of trajectories in the molecular dynamics of large systems. In this case, an acceptable exact approximation of the entire ultra-multidimensional and energetically accessible region of the potential energy surface is not required (in reality, this is impossible), and it is sufficient to approximate only a relatively small, smooth and dynamically accessible regions of the PES by calibrating the potentials.

As we know the problem of the topography of the ultra-multidimensional energy surface of polymers and biopolymers has several complex aspects.

To date, the physical theory in this area has been practically limited by phenomenological concepts of a relatively smooth structure of a multidimensional energy surface in the form of a potential funnel (the principle of minimal frustration of an energy funnel [14]). There are also no clear physical causes for which some linear polymers fold into unique spatial structures, whereas others (even those very similar in chemical composition) do not. Above, we briefly discussed why the problem of calculating the PES cannot be solved by summing the commonly used atom-atom potentials due to an uncontrollable error in summing a huge number of terms. From a purely technical standpoint, the calculation, for example, of a 100-dimensional surface with acceptable accuracy (even deviating from the accuracy of setting the potentials) with a step of, for example, 3.6° for each angle will require calculating 10^{200} numbers. This is unrealistic even with the most optimistic predictions about the development of supercomputer performance. At the same time, without understanding the basic principles of such multidimensional PES properties, it is almost impossible to further develop the physical theory of the folding into the unique spatial structures of biopolymers and their complexes, for example, in chromatin. In this situation, in our opinion, attention should be paid to establishing the analytical structure of this type of multidimensional surfaces taking into account the main features of the considered macromolecular objects.

It is known that valence bonds and valence angles have almost no contribution to the change in the spatial conformation of linear polymers and biopolymers, and the change in their spatial structure occurs due to rotations about single bonds. Thus, the topology of the configuration space of such polymers corresponds to the topology of a multidimensional torus, and the potential energy function that is defined on this torus must be represented as a multidimensional Fourier series. When considering the general properties of a multidimensional surface, we have to start from this analytical representation of a multidimensional PES and take into account various circumstances that

impose restrictions on the coefficients of the expansion of the PES in a Fourier series. Previously, we showed [20, 21] that certain restrictions on the expansion coefficients lead to such types of free energy surface topography that make it possible to interpret very different kind of facts observed during the folding of protein structures. In the second part of this article we showed how the symmetry of linear polymers with respect to the permutation of identical units limits the possible form of the expansion coefficients of the PES in the Fourier series. In the ideal variant of the homopolymer, the expansion coefficients are invariant with respect to permutation of harmonic numbers for different angular variables. The most natural invariants of harmonic vectors are the lengths and the sums of the components of these vectors. Above, we considered the dependences of the expansion coefficients on these invariants of the algebraic vectors of harmonic numbers. The considered effects of symmetry for homopolymer structures lead to the fact that the global minima of potential energy (if any) correspond to helical structures. When temperature effects are taken into account and states with minimum free energy are considered, the helical structures will be deformed in a certain way. While studying possible variants of the topography of multidimensional potential energy surfaces and free energy surfaces, we proceeded to the consideration of particular cases or models that have the required symmetry properties. The Gaussian dependence of the expansion coefficients on the length of the vector of harmonic numbers, adjusted for a possible shift of the distribution maximum by some vector in the space of harmonic numbers, was selected as the basic model. The Gaussian dependence describes the situation when the space of harmonic numbers contains a region of a certain volume that makes the main contribution to the Fourier series. Analytically, the Gaussian dependence is convenient, because it makes it possible to obtain compact equations from which the properties of the corresponding macromolecular structures can be clearly traced and can be compared with the available general experimental facts (for example, for the processes of protein folding and denaturation).

It was found that, within the Gaussian model, PES depends on two generalized variables (57) that have a simple geometric meaning in the space of torsion angles: the radius of the hypersphere around the global minimum point and the distance from the global minimum point to the hyperplane that is perpendicular to the vector drawn to the maximum point of the Gaussian distribution in the space of harmonic numbers. This successful parametrization of the PES in the framework of the Gaussian model give a chance to understand the possible causes that underlie the principle of minimum frustration of the energy funnel discussed above [14]. The resulting topography of the multidimensional PES satisfies this principle and, under certain conditions, provides an opportunity for fast folding into a unique spatial structure. On the other hand, the considered PES topographies generally show that the entire multidimensional configuration space of a polymer in the general case is divided into separate but very broad conformation regions that are separated from each other by sufficiently high potential barriers. As a result, the folding of a polymer chain can be very sensitive to the region of the initial configurations (conformations) of the chain, which is obtained, for example, during the synthesis of a polymer or biopolymer or its force denaturation. Such effects were observed earlier when comparing the results of protein refolding depending on the unfolding way of the polypeptide chain. With the force method of unfolding using atomic force microscopes, refolding to the native state was prevented [38, 39]. Possibly, the co-translational folding of the polypeptide chain [36] is associated with the possibility of harmonizing the configuration space regions, which, on the one hand, are optimal for folding to the global energy minimum and, on the other hand, are the result of a specific mechanism of polypeptide synthesis on the ribosome.

Accounting for entropy effects and the transition to the free energy surface generally retains the main features of the energy surface topography. In this case, an important new aspect, the volcano-type profile of the FES, appears, which is usually observed in folding experiments [9, 10]. When we proceed to PES, a new parameter also automatically appears, i.e., the characteristic temperature of a macromolecule, which is proportional to the specific gain in potential energy per monomer (or per angular variable) upon the transition to the global minimum point of PES. Here, we also have relatively simple analytical expressions that allow us to trace the possible effects in detail. In particular, we can determine the temperature at which the folded state becomes unstable. By selecting the denaturation temperature inherent in protein structures, it is possible to estimate the required values of the energy of interaction between monomer units, which is ~5 kcal/mol within the framework of the considered approach, or of the order of the hydrogen bond energy in an aqueous medium. Note that this estimate is obtained automatically as a result that is based on the topology of the configuration space of a linear polymer, there is certain symmetry with respect to the permutation of monomer units, the model hypothesis that the space of harmonic vectors contains a region that makes the main contribution to the Fourier series but without additional physicochemical considerations.

Obviously, not all linear polymers can form the unique spatial structures that are required for the functioning of living systems. The above considerations of the role of symmetry effects in the PES topography and dynamic ordering of the conformational dynamics in a viscous medium lead to the idea of the possible mechanisms of formation and selection

during the physicochemical evolution of macromolecules, which tend to form unique spatial structures. We proceed from the need for a prebiological stage of molecular evolution, when, as a result of chemical processes and physical laws, the macromolecules that stably fold into unique spatial structures that could later form supramolecular structures with certain functions were selected. We emphasize that we do not mean specific biopolymers in their modern form, the mechanisms of formation and selection of which are separate and not entirely clear, even in formulation. The previous stage of the formation of molecular "constructs" with a stable spatial structure and functionality from some primordial soup is one of the possible variants of the evolutionary transition to the simplest forms of macromolecular complexes with some features characteristic of living systems. The fundamental issue is whether there are (and under what conditions) objective physicochemical causes for the directed evolution of macromolecules towards the formation of macromolecules with a unique spatial structure? We propose that this issue should be considered from the standpoint of the evolution of the energy landscapes of linear polymers and the dynamics of folding in the direction of unique spatial structures, i.e., spatial structures with a single global minimum that can be hit in a reasonable time and under achievable conditions. In this regard, above we drew attention to the possible effect of the combined action of two types of symmetry. The first is the formation of an energy landscape with many energy funnels according to Eq. (74). This type of FES may form, as shown above, in the case of a linear homopolymer or a polymer with similar side groups on the monomer units. For folding such a polymer into a unique spatial structure, it is necessary that the representative point be in the region of attraction of the central energy funnel. The retention of a representative point in the region of attraction of the central funnel can be facilitated by the effects of partial ordering of conformational motions in a viscous medium. This article presents two adjacent formulas (Eq. (76) and a repetition of Eq. (24.1)) showing how the conditions imposed on linear combinations of angular velocities of turns around torsion angles can be corresponded as a result of the effects of the two discussed types of symmetry. Note that the effects of viscosity in conformational dynamics, which are described in [19] and lead to an approximately equal distribution of the energy dissipation rates (as well as the potential energy change rates) over the chain nodes, make an additional contribution to the dynamic ordering of the conformational degrees of freedom of the macromolecule and make the rules for the motion of a representative point along multidimensional energy surfaces more stringent. Thus, within theoretical concepts developed, we see certain physical prerequisites for directed physicochemical evolution towards the formation of polymeric macromolecules with unique spatial structures.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest. The author declares that he has no conflicts of interest.

Statement on the welfare of humans or animals. This article does not contain any studies involving animals performed by the author.

REFERENCES

1. I. M. Lifshits, A. Yu. Grosberg, and A. R. Khokhlov, Sov. Phys. Usp. **22**, 123 (1979).
2. A. Y. Grosberg and A. R. Khokhlov, *Giant Molecules. Here, There, and Everywhere*, 2nd ed. (World Scientific, Singapore, 2011).
3. H. Frauenfelder, *The Physics of Proteins: An Introduction to Biological Physics and Molecular Biophysics* (Springer, New York, 2010).
4. R. L. Mossbauer, Hyperfine Interact., **33**, 199 (1987).
5. K. V. Shaitan, A. B. Rubin, Mol. Biol. (Moscow) **14**, 1046 (1980).
6. K. V. Shaitan and M. G. Mikhailyuk, Khim. Fiz. **20** (2), 3 (2001).
7. K. V. Shaitan, Biophysics **39**, 993 (1994).
8. K. V. Shaitan, in *Stochastic Dynamics of Reacting Biomolecules*, Ed. by W. Ebeling, L. Schimansky-Gefer, and Y. M. Romanovsky (World Scientific, Singapore, 2003), pp. 283−308.
9. K. A. Dill and J. L. MacCallum, Science **338**, 1042 (2012).
10. A. V. Finkel'shtein, Usp. Biol. Khim. **58**, 7 (2018).
11. A. V. Finkel'shtein and O. V. Ptitsyn, *Protein Physics* (Knizhnyi Dom Univ., Moscow, 2012) [in Russian].
12. W. A. Eaton, J. Phys. Chem. B. **125**, 3452 (2021).
13. Yu. F. Krupyanskii and V. I. Gol'danskii, Phys.—Usp. **45**, 1131 (2002).
14. J. N. Onuchic and P. G. Wolynes, Curr. Opin. Struct. Biol. **14**, 70 (2004).
15. S. Piana, J. L. Klepeis, and D. E. Shaw, Curr. Opin. Struct. Biol. **24**, 98 (2014).
16. K. V. Shaitan, M. A. Lozhnikov, and G. M. Kobel'kov, Biophysics **62**, 182 (2017).

17. K. V. Shaitan, F. Yu. Popelenskii, and G. A. Armeev, Biophysics **62**, 348 (2017).

18. A. A. Erendzhenova, G. A. Armeev, and K. V. Shaitan, Biophysics **65**, 731 (2020).

19. K. V. Shaitan, Biophysics **63**, 1 (2018).

20. K. V. Shaitan, Biophysics **63**, 485 (2018).

21. K. V. Shaitan, Biophysics **63**, 675 (2018).

22. K. V. Shaitan, Biophysics **63**, 847 (2018).

23. J. P. Elliott and P. G. Dawber, *Symmetry in Physics: Principles and Simple Applications* (Oxford Univ. Press, New York, 1979; Mir, Moscow, 1983), Vol. 1.

24. L. D. Landau and E. M. Lifshits, *Mechanics* (Nauka, Moscow, 1988) [in Russian].

25. C. V. Heer, *Statistical Mechanics, Kinetic Theory, and Stochastic Processes* (Academic, New York, 1972; Mir, Moscow, 1976).

26. M. V. Vol'kenshtein, *Configuration Statistics of Polymer Chains* (Akad. Nauk SSSR, Moscow, 1959) [in Russian].

27. A. T. Fomenko, *Visual Geometry and Topology. Mathematical Images in the Real World*, 3d ed. (URSS, Moscow, 2018) [in Russian].

28. A. Zygmund, *Trigonometric Series* (Warsaw, 1935; Mir, Moscow, 1965), Vol. 2.

29. J. W. Milnor, *Morse Theory* (Princeton Univ. Press, Princeton, 1963; URSS, Moscow, 2011).

30. D. J. Wales, *Energy Landscapes* (Cambridge Univ. Press, Cambridge, 2003).

31. F. H. Stillinger, *Energy Landscapes, Inherent Structures, and Condensed-Matter Phenomena* (Princeton Univ. Press, Princeton, 2016).

32. V. V. Marchenkov and G. V. Semisotnov, Int. J. Mol. Sci. **10**, 2066 (2009).

33. J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale Univ. Press, London, 1902; Nauka, Moscow, 1982).

34. K. V. Shaitan and I. V. Fedik, Biophysics **60**, 335 (2015).

35. K. V. Shaitan and I. V. Fedik, Biophysics **53**, 44 (2008).

36. G. Kramer, D. Boehringer, N. Ban, and B. Bukau, Nat. Struct. Mol. Biol. **16**, 589, (2009).

37. A. S. Lemak and N. K. Balabaev, Mol. Simul. **15**, 223 (1995).

38. D. T. Edwards, M.-A. LeBlanc, and T. T. Perkins, Proc. Natl. Acad. Sci. U. S. A. **118**, e2015728118 (2021).

39. A. Bordgia, P. M. Williams, and J. Clarke, Annu. Rev. Biochem. **77**, 101 (2008).

Translated by M. Batrukova