

Role of evolution in shaping the unique structure of polypeptide chains

S. V. Panyukov

P. N. Lebedev Physics Institute, Academy of Sciences of the USSR, 117924, Moscow

(Submitted 17 June 1991)

Pis'ma Zh. Eksp. Teor. Fiz. **54**, No. 2, 110–112 (25 July 1991)

Evolutionary selection leads to the formation of proteins which have an unambiguously determined tertiary structure.

The unique three-dimensional folding of proteins (the tertiary structure) remains an intriguing unresolved problem in biophysics. Garel and Orland¹ have proposed a realization of Hopfield's model² of an associative memory, in which a distinct set of interactions v_{ij} between units i and j would correspond to each configuration (i.e., to each sequence of units) of the molecule. As Shakhnovich and Gutin³ have asserted, a heteropolymer of this general type should go into a frozen spin-glass phase as the temperature is lowered. Such a phase would have a memory: The state of the molecule would be determined not only by its configuration but also by its entire history.

In this letter we show that a "native" state with a fixed chain configuration can form for only those heteropolymers which have passed through the sieve of evolutionary selection, which has left us with only those molecules which have a minimum value of the energy $F_0\{v_{ij}\}$. In contrast with a spin-glass phase, such molecules would "recall" only one "pattern" as the temperature was lowered. This point would be of fundamental importance to their biological functioning.

The distribution of $\{v_{ij}\}$ in the "primordial soup" can be assumed Gaussian with a variance³ $\langle v_{ij}^2 \rangle_0 = v_0^2$. We describe the survival of molecules in the evolution process by the distribution function

$$w\{v_{ij}\} = \text{const} \cdot \exp(-\beta_0 F_0\{v_{ij}\}), \quad (1)$$

which is characterized by the parameter β_0^{-1} . In the initial stages of evolution, this parameter is the same as the temperature T_0 . Later on, because of a possible kinetic selection rule banning certain configurations of the molecules, the effective temperature becomes $\beta_0^{-1} < T_0$.

By virtue of the self-averaging nature of the free energy of a molecule, we can write

$$F = \langle F\{v_{ij}\} \rangle \equiv \langle w\{v_{ij}\} F\{v_{ij}\} \rangle, \quad (2)$$

$$\exp(-\beta_k F_k\{v_{ij}\}) = \prod_i \int d\vec{x}_i g(\vec{x}_i - \vec{x}_{i+1}) \exp[-\beta_k \sum_{i,j} v_{ij}(\vec{x}_i - \vec{x}_j)], \quad (3)$$

$$v_{ij}(\vec{x}) = v_{ij} \Delta(\vec{x}) + \bar{v}(\vec{x}),$$

where $T = \beta_k^{-1}$ and $F = F_k$ ($k \neq 0$) are the temperature and free energy under the

experimental conditions. The function g , which describes the connectedness of the units in the chain, and also the function Δ fall off over a length scale equal to the unit size of a unit, a .

By analogy with Ref. 3, we use a replica method to find an average over v_{ij} . Using a Hubbard–Stratonovich transformation, we introduce the order parameter

$$Q^{kl}(\vec{x}, \vec{x}') = \left\langle \sum_i \delta(\vec{x} - \vec{x}_i^k) \delta(\vec{x}' - \vec{x}_i^l) \right\rangle, \quad k \neq l. \quad (4)$$

This transformation differs from that of Ref. 3 in that it describes the correlation in the chain paths not only in different replicas $k = 1, \dots, m$ of the system under the experimental conditions but also in the path of the $k = 0$ chain, formed by evolution matrix $w(1)$, in terms of the sequence in which the units of the “pattern” alternate. We will show that the existence of this standard pattern of the tertiary structure makes it possible to realize the basic ideas of a Hopfield associative memory.¹

In the globular regime of interest here, the quantities Q^{kl} should be found from the condition for a minimum⁴ of free energy (2):

$$mF/T = \sum_{k \neq l} \int d\vec{x} d\vec{x}' \beta_k \beta_l (\hat{Q}^{kl})^2 + \sum_k \beta_k E(\rho_k) - \ln Z\{\hat{Q}\},$$

$$Z\{\hat{Q}\} = \prod_{i,k} \int d\vec{x}_i^k g(\vec{x}_i^k - \vec{x}_{i+1}^k) e^{-\sum_i H(\vec{x}_i)}, \quad (5)$$

where $E(\rho)$ is the energy of the interaction of \bar{v} units of the chain, ρ is the density of these units, and the self-consistent field H which is exerted on these units by the tertiary structure is

$$H(\vec{x}) = - \sum_{k \neq l} \beta_k \beta_l \int d\vec{x}' \Delta(\vec{x}^k - \vec{x}') \hat{Q}^{kl}(\vec{x}', \vec{x}'),$$

$$\hat{Q}^{kl}(\vec{x}, \vec{x}') = \int d\vec{x}'' Q^{kl}(\vec{x}, \vec{x}'') \Delta(\vec{x}'' - \vec{x}'). \quad (6)$$

To calculate $\ln Z$ we use Feynman’s variational principle

$$\ln Z \geq \ln Z_{\text{tr}} + \left\langle \sum_i [H_{\text{tr}}(\vec{x}_i) - H(\vec{x}_i)] \right\rangle_{\text{tr}}. \quad (7)$$

For Gaussian functions g and Δ , we choose the field of a potential well which confines the units of the chain in the tertiary structure as follows:

$$H_{\text{tr}}(\vec{x}) = \sum_{k \neq 0} q^{0k} (\vec{x}^k - \lambda \vec{x}^0)^2 + \sum_{k \neq l} q^{kl} (\vec{x}^k - \vec{x}^l)^2, \quad (8)$$

where λ is the degree of extension (swelling) of the molecule. The quantities q^{kl} , which are measures of the amplitude of the fluctuations in the chain with respect to its

average path, are to be found by maximizing the right-hand side of (7). The Gaussian integrals in (7) can be evaluated easily.

The equations for Q have the trivial solution $Q^{kl} = \rho_k / V_l$ with $q^{kl} = 0$, which describes an ordinary globule of volume $V_k = V_0 \lambda^3$ in the absence of a self-organized structure. The native state of the molecule is described by the nontrivial solution with the Fourier components

$$Q^{kl}(\vec{p}) = \rho_k \exp(-\sigma^{kl} \vec{p}^2), \quad k < l. \quad (9)$$

The maximum of F is reached on a matrix σ which is symmetric with respect to the replicas $k = 1, \dots, m$. This situation corresponds to the memory of a single pattern per molecule. A solution of this sort exists only under the condition

$$\kappa = (\beta_0 v_0)^2 \rho_0 a^3 > 1.$$

Consequently, a statistically reliable native state can arise only in the late stages of the evolution, with $\beta_0 > \beta_0^c$.

A first-order phase transition (a denaturing) occurs from the native state to a molten globular state with the same density ρ at the temperature

$$T_c = \kappa / \beta_0 (\tilde{a} + \tilde{b} \lambda^2), \quad \tilde{a}, \tilde{b} \sim 1.$$

This transition has a significant latent heat (on the order of T per unit). The length scale R of the fluctuations in the path of the chain ($R^2 = \sigma \approx a^2 \beta_0 T / \kappa$) at the transition point $T = T_c$ is equal in order of magnitude to a and decreases with decreasing temperature. The fluctuations thus do not disrupt the unambiguous definiteness of the tertiary structure, and the phase transition may be thought of as a melting of this structure.

The evolutionary distribution function which we have selected (1), leads to the distribution function

$$p(\epsilon) = \text{const} \cdot \exp(-\beta_0 \epsilon - \epsilon^2 / 2v_0^2)$$

of the interaction energies ϵ of the units in the tertiary structure which are spatially closest together. The Boltzmann factor $\exp(-\beta_0 \epsilon)$, which screens out configurations that are unfavorable from the energy standpoint, becomes the primary factor at $\beta_0 v_0 \gg 1$. A statistical dependence of this sort has been seen experimentally with an effective temperature $\beta_0^{-1} \approx 200$ K.

A more detailed description of the unique native structure of proteins could be found by solving the corresponding kinetic problem of determining the distribution function $w\{v_{ij}\}$. As we know from analysis of the Hopfield model, the distortions of the stored pattern in comparison with the case in (1) (because of a kinetic depletion of configurational selection) can be significant without disrupting the single-valued nature of the recognition of the tertiary structure.

¹T. Garel and H. Orland, Europhys. Lett. **6**, 307 (1988).

²J. J. Hopfield, D. J. Feinstein, and R. G. Palmer, Nature **304**, 158 (1983).

³E. I. Shakhnovich and A. M. Gutin, Biophysical Chemistry **34**, 187 (1989).

⁴J. R. L. de Almeida and D. J. Thouless, J. Physique A **11**, 983 (1978).

Translated by D. Parsons