FACTA UNIVERSITATIS
Series: **Physics, Chemistry and Technology** Vol. 14, N° 3, Special Issue, 2016, pp. 197 – 214
DOI: 10.2298/FUPCT1603197A

PROTEIN ULTRAMETRICITY AND DESIGNING OF SYNTHETIC MOLECULAR MACHINES

UDC 547.963:577.32

Vladik A. Avetisov[†]

The Semenov Institute of Chemical Physics,
Russian Academy of Sciences, Moscow, Russia

Abstract. *The idea of protein ultrametricity proposed nearly 35 years ago pertaining to CO binding by myoglobin is still questionable. In this overview, the contradicting attempts to describe the CO-rebinding kinetics in the framework of familiar approaches are discussed together with the ultrametric description. The overview is extended by new angle devoted to designing synthetic molecular machines using hierarchy and self-similarity as the design principles.*

Key words: *protein ultrametricity, molecular machines, hierarchy, self-similarity*

1. Introduction

Ultrametrics was first applied as a physical concept to statistical theory of spin glasses ([20], [22]). It is believed that the ground state of a spin-glass system is reached via the tree-like splitting of a phase space into hierarchically embedded basins of states. The branching points on the tree determine dissimilarities of the basins, so the metrics on the state space obey the strong triangle inequality, i.e., the state space is an ultrametric space.

Shortly thereafter, a similar idea was voiced with respect to proteins ([1],[13]). As applied to proteins, ultrametrics presumes that the local minima of the protein-energy landscape are clustered into hierarchically embedded basins of minima: Each large basin consists of smaller basins, each of these basins consists of even smaller ones, and so on. It is assumed that the energy barriers separating the basins are

Received April 2nd, 2016; accepted June 17th, 2016.
[†]**E-mail:** vladik.avetisov@gmail.com

arranged in such a way that the equilibration within any basin is attained for a much shorter period of time than the time necessary to leave the basin. When there is a hierarchy of barriers separating the basins, only the highest barrier between any given two states specifies the transition. As a result, the kinetic distances between the states measured by the transition times obey the strong triangle inequality, i.e., the protein states associated with local energy minima on the energy landscape form an ultrametric space.

Despite the fact that ultrametrics appeared in spin glasses and proteins almost simultaneously, proteins are very far from disordered systems based on their ability to execute exact operations with individual atoms or molecules. Protein functionality has long been a clash of two mutually excluding interpretations: A historical primer view refers to biochemical catalysis and a more contemporary paradigm of molecular machines (see, for example, [11]). Under the primer view, proteins reduce the activation barrier and catalyze the reaction. According to this viewpoint, a catalytic center (the protein active site) is in the focus, and the majority part of a protein molecule surrounding the active site is considered to be a rigid matrix that ensures selective binding via key-lock complementarity. In short, very specific tertiary (spatial) structures of proteins ordained by the primary structures are the most important for protein functioning. Therefore, the primer view reflects above all the central dogma of biology widely referred as the idea that the genotype determines the phenotype.

The paradigm of molecular machines presents an opposing viewpoint. Now, protein dynamics play a key role. It is believed that specific excitations of the active site by chemical binding, charge transfer, etc., launch the multi-scale rearrangements of a protein during which the excitations are transferred from fast degrees of freedom to the slowest quasi-mechanical motion of particular fragments performing the protein function. It is this feature that imparts protein functioning upon the character of a machine action. In this view, protein conformational dynamics are of upmost importance.

Although it has long been proven that the catalytic interpretation leads to physically senseless estimations of the activation energy and entropy of the most biochemical reactions [11], the interpretation still survives.This survival is due not so much to the fact that a rather-simple idea is difficult to discard, but rather due to the lack of a description of protein dynamics acceptable over a wide range of time scales. Actually, the problem is easy to see from its formal statement. In general, if the energy landscape, $\Phi(r)$, is somehow specified on the states, $\{r\}$, of a system ($r \in \mathbb{R}^N$, where N is the number of degrees of freedom of the system), one can study the dynamics on the manifold $\Phi(r)$ using analytical, numerical, or computer simulation methods. There have been a variety of studies devoted to even relatively complicated molecular structures (see, for example, [9] and [25]). However, proteins are overly complicated for reconstructing entire functional cycles in such a way. Computer simulations provide either a highly resolved picture of protein dynamics in relatively small areas of conformational space or a coarse-grain image of the movements of large protein fragments. It seems impossible to reconstruct in

detail the entire process of transmission of local excitations at an active site into the directed quasi-mechanical movements of large structural fragments. Analytical modeling is a challenging problem here.

In any case, one needs to make some simplifications in order to describe the multi-scale protein dynamics. In this respect, it seems that ultrametric random processes, which are inherently multi-scale, open up new perspectives in justification of the molecular machine concept. This idea is summarized in Section 2. in which we demonstrate considerable progress in the ultrametric description of the CO-binding to myoglobin. On the other hand, advances in ultrametric modeling of protein dynamics provoke a new angle for designing synthetic molecular machines based on hierarchy and self-similarity as the design principles. Section 3. presents recent simulations supporting this idea.

2. Protein ultrametricity and CO-binding by myoglobin

2.1. Experiments

The studies of CO-binding by myoglobin ([1], [13], [23]) occupy a special place among attempts to clarify how protein dynamics control protein function. The experiments of these authors were rather simple. At given temperature, T , a sample with myoglobin initially bound to CO was irradiated by a laser pulse. In doing so, one breaks a chemical bond with CO in a part of the myoglobin molecules in the sample. Immediately after, the CO-rebinding kinetics were monitored, specifically the relative concentration, $n(t, T)$, of the myoglobin molecules still unbound at an instant t . These data were measured over the short time-scale resolution of approximately 10^{-7} s up to long time scales of approximately tens of seconds and more. The temperature range was also very wide and varied from room temperature to deeply frozen samples at 60K and even below.

Qualitatively, the overall kinetic picture looks as follows (for kinetic data, see [1] and [23]). First, there was a high-temperature range ($300 \div 200$K) and a low-temperature range ($190 \div 60$K) over which the rebinding kinetics differed significantly. The kinetic curves at high temperatures have two characteristic segments: One relates to power-law decay on intermediate time scales, and one relates to exponential decay on final time scales. As the temperature decreases, the power-law decay extends to a larger and larger part of the time window and covers the entire window at the bottom of the high-temperature range at approximately 200 K. The binding rate increases on intermediate time scales and decreases on long time scales. In other words, on long times CO-rebinding exhibits ordinary first-order chemical kinetics; on the intermediate time scales the same reaction is anomalous with respect to both the kinetics and the temperature behavior. In the low-temperature range (19060 K), the picture is not so enigmatic. The rebinding kinetics are non-exponential and ordinarily depend on the temperature, i.e., the reaction rate decreases with decreasing temperature.

In fact, not all of the kinetic features noted above correspond with the protein dynamics. When temperatures are sufficiently high, CO molecules after laser pulse may either remain inside the protein or leave the protein [23]. In the last case, the slow penetration of CO molecules through the globule to the binding site limits the rebinding rate, and the kinetics are exponential. In contrast, the rebinding of those CO molecules that remain inside the protein is regulated by protein dynamics. Therefore, at high temperatures only the power-law segments of the kinetic curves are of interest. At low temperatures, an exit channel is closed so all of the CO molecules remain inside the proteins. At low temperatures, the kinetics are of interest over the entire time window.

In summary, when considering only the rebinding that is regulated by protein dynamics, the rebinding kinetics are always non-exponential. However, the reaction rate anomalously depends on temperature in the high-temperature range and normally depends on temperature in the low-temperature range.

Quantitatively, it turns out that a power-law decay of the form

$$n(t,T) \approx \left(\frac{t}{\tau}\right)^{-\left(1-\frac{T}{T_0}\right)} \tag{1}$$

describes the kinetics well in the high-temperatures regime at $t \gg \tau_0$, $200 < T < T_0 \approx 300\text{K}$ [23]; the power-law decay of the form

$$n(t,T) \approx \left(\frac{t}{\tau_{1/2}(T)}\right)^{-\frac{T}{T_0}} \tag{2}$$

describes the kinetics well in the low-temperature regime at $t > \tau_{1/2}(T), T < 180\text{K}$ ([27]). In the equations listed above, T_0, and τ are scaling parameters, respectively, and $\tau_{1/2}(T)$ is the half-life time of unbound proteins.

2.2. Interpretations

An initial interpretation of the CO-rebinding kinetics was suggested by the authors of the experiments [1], [13], and [23] in spirit of the catalytic concept. They stated that the proteins in a sample are frozen in individual conformational substates so the kinetics are non-exponential due to slightly different activation barriers. According to this viewpoint, a multi-mode kinetic model was proposed in the form:

$$n(t,T) = \int g(H) \exp\{-k(H,T)\} dH, \tag{3}$$

where $g(H)$ is the distribution of activation barriers H, and $k(H,T)$ is the binding rate constant. The authors of the experiment believed that frozen proteins were much closer to the glass-like systems than to the native proteins. Such a seemingly natural interpretation, however, led to difficulties; to obtain the power-law decays (1) and (2), the authors were forced to admit very specific distributions of activation

barriers incompatible at high temperatures and low temperatures that inexplicably varied with temperature and time [23].

Together with all this, they surprisingly declared that the rebinding kinetics suggest the ultrametricity of the protein conformational substates in the sense of a hierarchy of basins of local minima on the protein-energy landscape [13]. As a result, the idea of protein ultrametricity has been ambiguous from the very beginning. In contrast to the ultrametricity of glasses, the authors declared this idea without arguments how exactly ultrametricity relates to the rebinding kinetics; moreover, they insisted that the interpretation of the experiments had nothing to do with ultrametricity.

Slightly later, an opposite interpretation of CO binding in the spirit of the molecular machine concept was proposed in [26]. The authors of this interpretation argued that precisely at low temperatures, i.e., exactly where the assumption about frozen conformations looks more believable, just the protein dynamics limit the rebinding kinetics. According to this interpretation, the following model was proposed:

$$n(t) = \left\langle \exp\left\{ - \int_0^t k(x(t'))dt' \right\} \right\rangle, \tag{4}$$

where $k(x(t))$ is the binding rate constant controlled by the protein dynamics presented by a random process $x(t) \geq 0$ and $\langle \cdot \rangle$ is the average over the trajectories $x(t)$. The control was specified as follows: $k(x(t)) = k_0$ if $0 \leq x(t) \leq 1$, and $k(x(t)) = 0$ otherwise, i.e., myoglobin can bind CO only when the protein gets into a specific set of conformational substates given by the interval $[0, 1]$.

Note that the model (4) also has nothing to do with ultrametricity. However, in contrast to the primer interpretation (3), it implies a particular relationship between the binding kinetics and the protein dynamics. Despite the fact that there is no explicit form of such a relationship in the model (4), the authors [26] attempted to specify the characteristics of the process $x(t)$ that would be compatible with the observable kinetic features. They argued that the process $x(t)$ was relevant to the established low-temperature kinetics (2) if the mean number, $m(t, T)$, of getting into the interval $[0, 1]$ increases with time as $m(t, T) \sim t^{T/T_0}$. It is easy to see, however, that in this case the model (4) directly contradicts the anomalous temperature dependence at high temperatures. In other words, if the model (4) correctly describes the rebinding kinetics at low temperatures, it does not work at high temperatures, and vice versa.

In summary, a contradictory situation arose around the interpretations of the CO rebinding experiments. On the one hand, it remained unclear whether the protein structure or the protein dynamics were more important for protein functioning. On the other hand, it appeared that the native proteins and frozen proteins physically differed so much that it was hardly possible to expect the same behavior at high and low temperatures. In terms of protein ultrametricity, this idea was actually not discussed. Moreover, the belief spread that the CO-rebinding kinetics permitted different interpretations having nothing to do with ultrametricity. In

fact, a consistent picture of the CO rebinding kinetics at all temperatures can be constructed based namely on protein ultrametricity.

2.3. Ultrametric models

In general, the ultrametric description of CO rebinding has been developed in a series of works based on the interpretation (4) (for the original studies, see [3], [4], and [5]; for a recent review, see [6]). The only difference is that the protein states are described by an ultrametric space. As a result, the ultrametric diffusion models the protein dynamics.

Let us start with a scheme of the CO rebinding: Mb-CO→Mb*→ ... →Mb_1+ CO→Mb-CO. In the scheme, Mb* denotes a set of states in which myoglobin molecules turn out immediately after breaking the chemical bonds with CO; Mb_1 denotes a set of states to which unbound proteins relax and in which they rebind CO. The quantity measured in the experiments is a fraction $n(t,T)$ of proteins remained unbound in a sample at a temperature T at an instant t . Note that the measured quantity $n(t,T)$ takes into account the proteins found in any conformational substates [Mb* → ... → Mb_1] ranging from the initial states Mb* up to the final states Mb_1. In the ultrametric description, an ultrametric ball $B_r \subset \mathbb{Q}_p$ of sufficiently large radius p^r, $(r \gg 1)$ describes all of these states, and an ultrametric ball $Z_p \subset B_r$ with a unit radius represents the chemically active states Mb_1 . In these terms, the initial states Mb* are distributed on a compact support $M \subset B_r$, $M \cap Z_p = \varnothing$.

According to the main idea, the following p-adic equation of the reaction-diffusion type describes the CO binding

$$\frac{\partial f(x,t)}{\partial t} = \int_{B_r} w(x|y)\,[f(y,t) - f(x,t)]\,d_p y - \lambda\Omega(|x|_p) f(x,t), \tag{5}$$

where $f(x,t)$ is the transition probability density of ultrametric diffusion, $d_p x$ is the integration measure on the field of p-adic numbers, Q_p , λ is the binding constant, and $\Omega(|x|_p)$ specifies the binding area ($\Omega(|x|_p) = 1$ if $|x|_p \le 1$, and $\Omega(|x|_p) = 0$ otherwise). The physical meaning of (5) is rather simple: an unbound protein diffuses over an ultrametric space of states, B_r, and binds CO only when the protein gets to the binding area, Z_p . The measured quantity in model (5) is $n(t) = \int_{B_r} f(x,t) d_p x$.

The novelty of the ultrametric model (5) consists of the fact that it introduces a direct relationship between the protein ultrametric dynamics and the rebinding kinetics. Specifically, the transition rates, $w(x|y)$, in (5) depend on the ultrametric distances $|x-y|_p$ as $w(x|y) = |x-y|_p^{-(\alpha+1)}$, where $\alpha \sim T^{-1}$. Such a form of transition rates explicitly assumes that the activation barriers separating the basins of local minima grow linearly with hierarchy level.

We note that the initial distribution may also be important for the kinetic differences at high and low temperatures. Indeed, the high-temperature kinetics (1)

may be related to a long-time asymptotic of the solution of (5), and the initial distribution shape is insignificant for such estimations. However, the low-temperature kinetics (2) may relate to the intermediate behavior of the same solution of (5), and for intermediate kinetics the initial distribution shape may be important. One can accordingly note that the breaking of a chemical bond at the binding site does not change noticeably the protein-energy landscape. Therefore, keeping the consideration (5) self-consistent, one can suggest that an initial distribution $f(x,0)$ be generated via ultrametric diffusion itself. Specifically, the initial distribution can be defined as follows:

$$f(x,0) = N \sum_{\gamma=\gamma_{\min}+1}^{\gamma_{\max}} \Omega(|x|_p = p^{\gamma}) p^{-c|\gamma-m|} p^{-\gamma}, \tag{6}$$

were N is the normalization coefficient. The initial distribution (7) has descending wings whose shapes are typical of an ultrametric diffusion distribution (see, for example, [3], [4]) with the maximum at the p-adic distance p^m from the binding area Z_p. Figure 1 shows the kinetic curves $n(t)$ calculated numerically for the model (5) with the initial conditions given in (6) (for the analytical solutions of p-adic equations of the reaction-diffusion type, see [10]). One can see that both the high-temperature kinetics ($\alpha < 2$) and the low-temperature kinetics ($\alpha \geq 2$) correspond to the power-law decays (1) and (2).

Figure 2 clearly shows that the anomalous temperature dependence changes to a normal temperature dependence as the temperature decreases. Notice that a single parameter $\alpha \sim T^{-1}$ parameterizes the change. It is interesting to note that the temperature dependence of the reaction rate changes to an opposite sense over a narrow range of temperatures (close to $\alpha = 2$). Such critical behavior is entirely consistent with the fact that myoglobin experiences the glassy transition resulting in a sharp decrease in protein-fluctuation mobility exactly at the border of high-temperature and low-temperature ranges.

In addition, it should be emphasize that the very specific features of spectral diffusion in deeply frozen proteins are described based exactly on the same ultrametric description of protein dynamics [2]. It is certainly a surprise that the protein dynamics demonstrate the universality over an extremely wide range of temperatures ranging from room temperature up to deeply frozen states.

2.4. Comments

As far as the barriers separating the basins specify the transition rates, the function $w(|x-y|_p)$ represents the protein-energy landscape. The transition function in the form $w(|x-y|_p) = |x-y|_p^{-(\alpha+1)}$ corresponds to the energy landscape with self-similar hierarchy of basins and barriers. Therefore, the ultrametric model (5) unambiguously assumes the self-similarity of the protein-energy landscape. The analytical arguments supporting this choice are outlined in [5]. It seems that there are profound physical reasons behind this fact: The self-similarity of the protein-energy

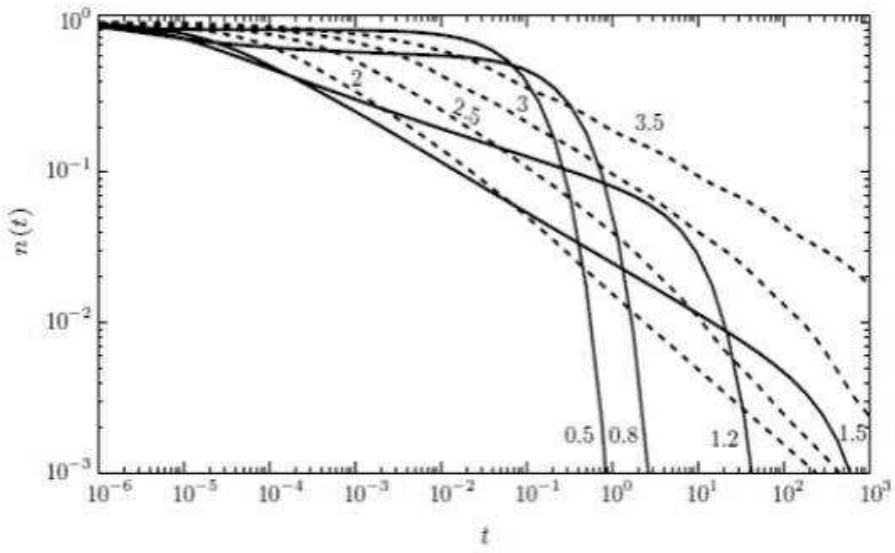

FIGURE 1: Kinetic curves $n(t)$ calculated following the ultrametric model (5)–(7) for the parameters $\tau = 10^{-7}$ s, $p = 2$, $r = 20$, $\gamma_{\min} = 5$, $\gamma_{\max} = 15$, $m = 10$, $\lambda = 10$; the values of α are shown above the curves. The high-temperature regime corresponds to the values $\alpha = 0.5 \div 1.5$ (solid lines). The low-temperature regime corresponds to the values $\alpha = 2.0 \div 3.5$ (dashed lines).

landscape underlies protein functionally over a wide range of temperatures. This finding implies that, when studying frozen proteins, even at cryogenic temperatures, one can obtain information relevant to native proteins. The wide-spread belief that the dynamics of deeply frozen proteins have nothing to do with the dynamics of native proteins is accordingly misguided. Furthermore, hierarchy and self-similarity may turn out to be guiding principles in the designing of macromolecular structures equipped with molecular machine functionality. The next section presents this idea.

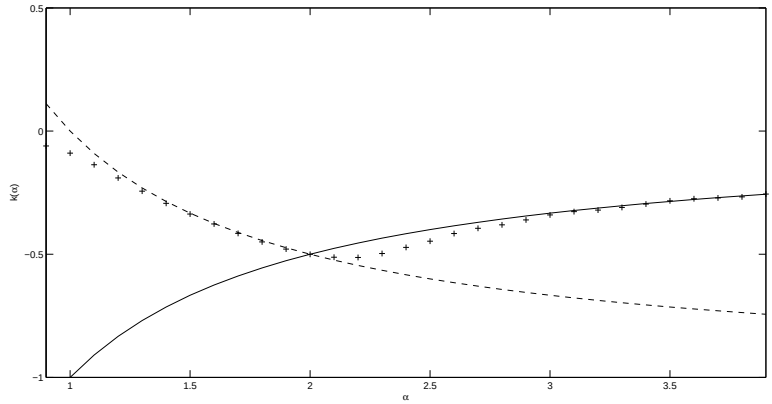

FIGURE 2: Dependence of the power–law exponent $k(\alpha)$. The solid and dashed curves relate to the expressions (1) and (2), respectively, and the dotted curve relates to the ultrametric model (5)–(7).

3. Ultrametric concept of designing molecular machines

Molecular machines are commonly attributed to complex macromolecular structures able to convert the perturbations of fast degrees of freedom into slow quasi-mechanical motion along a low-dimensional path. They operate accurately with individual atoms and molecules, they carry out assembling and disassembling of biopolymers, they move sub-cellular units along particular tracks in a cell, etc. Is it possible to design similarly operating synthetic macromolecules? This challenging question addresses the exiting prospects of algorithmic chemistry and waste-free technologies. Beyond chemistry, the question pertains to the core problem of the origin of life consisting of the fact that even the simplest version of biological reproduction based on translation and transcription remains hopelessly complex for its random appearance in abiogenic conditions ([12], [16], [7], [17], [19], [21]). In fact, to produce biological macromolecules, it is necessary to have the molecular machines able to accurately assemble at the atomic level. However, it is still unclear whether molecular machines may exist in abiogenic conditions, and what the primary machine would be. To make progress in resolving these problems, we combine two ideas. One idea follows protein ultrametricity and suggests that hierarchy and self-similarity are the basis of molecular machine design. Another idea requires the conditions natural for a lifeless environment for forming the target structure. Together, these ideas are focused on so-called crumpled polymer globules.

3.1. Crumpled globules

It is well known that, at high temperatures (in a good solvent), a polymer of

N units is a strongly fluctuating coil without a well-defined thermodynamic state. However, at temperatures below the critical value θ (in a poor solvent), a polymer chain collapses into a weekly fluctuating drop-like globule of size $R_{gl} \sim N^{1/3}$ (see, for example, [14]). An ordinary globule formed from a knotted chain with free ends all being subchains of length $s \leq N^{2/3}$ resembles mutually entangled Gaussian coils. For unknotted polymer rings, however, the globular structure is different [15]. Topological constraints inherent to a polymer ring result into so-called crumpled globule. The following imaginative hierarchical process elucidates the formation of a crumpled globule. Below the θ-point, there exists a relatively short length, g^* , such that the subchains of the order of g^* first collapse and form the crumpled units in a globular phase. Next, several crumpled units collapse and form the first-level folds; these folds then form the second-level folds, etc. This process continues up until the formation of the largest fold from the entire chain (see Figure 3a).

This recursive collapse functions mainly to illustrate the formation of hierarchically folded chains resembling weakly entangled globules on all scales $s > g^*$. Note that hierarchical collapse imposes constraints on chain entanglement on all scales.

3.2. Ultrametric description of a hierarchy of folds

Any crumpled unit in a hierarchy of folds can be parameterized by a sequence of indices specifying the set of embedded folds (e.g., to which particular first-level fold embedded into particular second-level fold, etc.) housing each crumpled unit. The terminal nodes (leaves) of a branching tree of folds encode the set of indices: Each leaf is specified by a unique path from the root to the leaf (see Figure 3b). Therefore, the tree boundary represents a space of states of the crumpled units in a hierarchy of folds. The number of downward branches of the tree sets a number of previous-level folds embedded into a fold of the next level. In a self-similar hierarchy of folds, the branching index p is fixed, and all γ–level folds consist of p^γ units (as shown in Figure 3b for $p = 2$).

Now, to describe how a polymer chain is superimposed on a hierarchy of folds, one can introduce two metrics specifying the distances between the crumple units. One distance is a conventional distance, s , between units measured along a chain. Another distance, d , specifies proximity of the states of units in a hierarchy of folds. Since the states correspond to the tree boundary, the metric d obeys the strong triangular inequality, i.e., the states of crumpled units in a hierarchy of folds form an ultrametric space. For a regularly branching tree of folds, the ultrametric distance d between any two units can be measured by the scale of the maximal fold separating the units. Note that crumpled units located nearby in real space may belong to different large fold and therefore be separated by a large ultrametric distance d.

In these terms, a hierarchically folded chain can be interpreted as a trajectory of a random process propagating with a polymer length s over an ultrametric space. The transition probability, $f(x, s|x_0, 0)$, of the random process specifies the probability of finding a unit s at the state x given the condition that the first unit of s-length

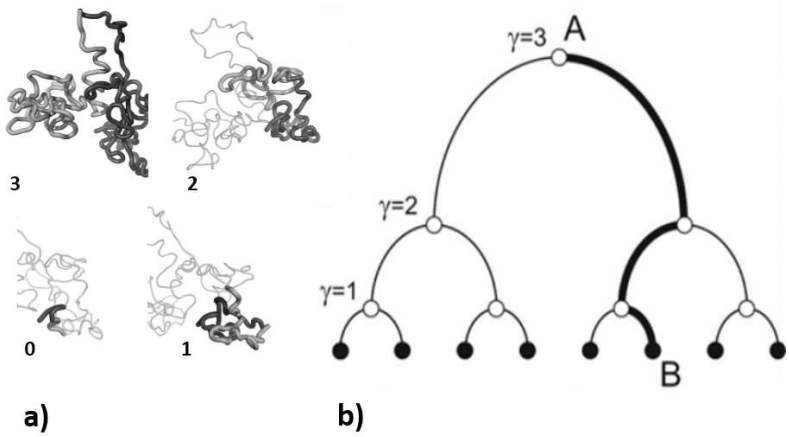

FIGURE 3: Hierarchy of folds of three levels. a) Representation of folds embedded into each other: 0 – crumpled unit, 1 – fist-level fold consisting from 2 crumpled units, 2 – second-level fold consisting from 2 first-level folds, and 3 – third-level fold consisting from 2 second-level folds. b) Tree-like presentation of the hierarchy of folds: The leaves on a tree form a space of states for the crumpled units. Each subtree of the tree corresponds to a particular fold. Each path AB on the tree from the root A to a leaf B specifies a set of embedded folds related to a particular state of a crumpled unit.

fragment is at the state x_0 . The transition probability function $f(x, s|x_0, 0)$ measures all of the possible conformations of hierarchically folded s-length fragments whose ends are in states x_0 and x , respectively. Implying conditional probability, we can write the transition probability $f(x, s|x_0, 0)$ in the abbreviated form $f(x, s)$.

Let the transition probability obey the master equation for the random work

$$\frac{\partial f(x,t)}{\partial s} = \sum_{x \neq y} w(x|y) f(y,s) - \sum_{x \neq y} w(y|x) f(x,s) \tag{7}$$

subject to the initial condition $f(x, 0) = \delta(x - x_0)$, where $w(x|y)$ is an element of the transition rates matrix $\mathbf{W}$ specifying the probability of jumping from state y to state x via unit step along the chain. Assuming $w(x|y) \sim p^{-\gamma(x,y)}$, where $p^{\gamma(x,y)}$ is the scale of a largest fold separating the states y and x, and taking into account that the state x is chosen in the $\gamma(x, y)$-level fold randomly, one can specify the matrix elements as $w(x, y) \sim p^{-2\gamma(x,y)}$.

Accordingly, the transition matrix $\mathbf{W}$ has a characteristic block-hierarchical structure related to the hierarchy of folds. For a regularly branching tree of folds, the hierarchy of blocks is also regular. In an irregular tree of folds, the scales of folds belonging to the same level may differ. In this case, the transition matrix

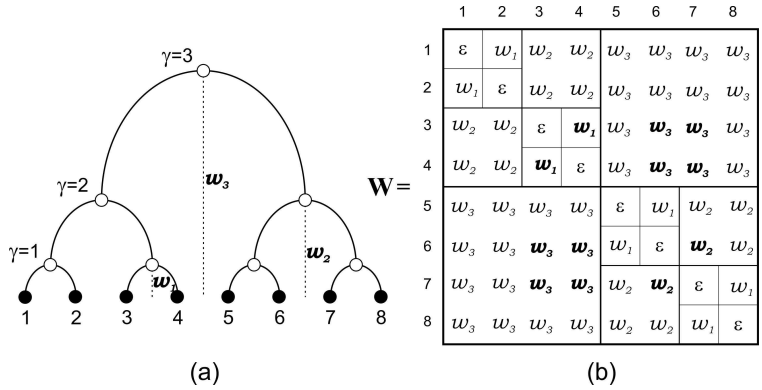

FIGURE 4: Block-hierarchical structure of a transition matrix ($p = 2$) (right) for random walk on the boundary of a regular tree of folds (left).

W will have again a block-hierarchical form, yet the matrix elements will vary in blocks even belonging to the same level of hierarchy.

The non-zero eigenvalues of the transition matrix are of primary importance: Each eigenvalue corresponds to a particular fold, so a fold is stable if its eigenvalue is negative. In the case of a regular tree of folds, the eigenvalues of are well known (see, for example, [3] and references therein:

$$\lambda_\gamma = -\frac{p - p^{-1}}{p - 1}\left(p^{-\gamma} - p^{-(\gamma_{max}+1)}\right) - p^{-(\gamma_{max}+1)}, \tag{8}$$

where each eigenvalue λ_γ is degenerated $p^{(\gamma_{max}-\gamma)}$ times, and $\gamma = 1, 2, \ldots, \gamma_{max}$ refers to the hierarchy levels. It is easy to see that all eigenvalues are negative. Therefore, the hierarchy of folds described by the equation (7) has a well-defined ground state.

Again, (7) implies the hierarchy of constraints, where each constraint stabilizes the folds of a particular scale. In this sense, the conditions for the eigenvalues to become positive with increasing temperature define the destruction of the folds. The destruction (melting) of folds can be illustrated by introducing a temperature-dependent repulsion between the folds, assuming that the repulsion can be described by an auxiliary block-hierarchical matrix as follows:

$$\frac{\partial f(x,t)}{\partial s} = \sum_{x \neq y} w(x|y) f(y,s) - (1+\tau) \sum_{x \neq y} w(y|x) f(x,s), \tag{9}$$

where $\tau = \frac{(T-\theta)}{\theta}$. Now, the eigenvalues are equal to $\lambda_\gamma^{(\tau)} = \lambda_\gamma + \tau$. The solutions of the equations $\lambda_\gamma^{(\tau)} = 0$ specify the critical temperatures $\theta < T_{\gamma_{max}} < T_{\gamma_{max}-1} < \ldots < T_1$ at which the folds are subsequently destroyed: At $T_{\gamma_{max}}$ the largest fold becomes unstable and splits into a set of $(\gamma_{max} - 1)$–level folds. These folds become

unstable at the next critical temperature $T_{\gamma_{max}-1} > T_{\gamma_{max}}$ and so on until the highest critical temperature, T_1 , is attained and the crumpled globule completely melts. Note that the melting of a regular hierarchy of folds is suggested to be a step-wise process due to well-distinguished folds; an irregular hierarchy of folds may melt almost continuously.

3.3. Designing of hierarchically folded samples on a computer

A polymer chain consisting of N units (beads) has been modeled using a standard set of potentials accounting for rigidity and volume interactions (see, for example, [18]). According to the ultrametric description noted above, the hierarchically folded samples were designed using a trial attractive interaction, U_γ , of a block-hierarchical form applied to a set $M \leq N$ of representative units. At the initial stages of the chain collapse, the neighboring representative units were attracted to each other to form the first-level folds. Then, the pairs of such folds formed the second-level folds, and so on. In general, the attraction of two γ–level folds with the formation of a fold of the next $(\gamma + 1)$ level was possible if the two folds were located close to one other. The spatial distance between two folds has been defined as the smallest distance between the units included in the folds. Furthermore, any fold can decay on sub-folds only if the upper-level fold is destroyed.

The collapse of relatively short chains with lengths of roughly 200 units has been performed using the Metropolis algorithm in a large box with open boundary conditions. When collapsing the chains and equilibrating the final conformations, we studied the dynamics of hierarchically folded samples using the elastic network technique. In an elastic network, a molecular structure is presented by a set of nodes $i = 1, \ldots, M$ initially located in a space at particular positions and elastic links between the nodes encoded in an adjacency matrix $\mathbf{A}$ with elements $a_{ij} = 1$ for a link between i and j , and $a_{ij} = 0$ otherwise. Therefore, the nodes are subject to elastic forces. In the overdamped limit, the velocity of each node is proportional to the sum of the applied elastic forces; hence, the relaxation of the molecular structure to equilibrium can be described by non-linear dynamic equations of the form:

$$\frac{d\mathbf{R}_i}{dt} = \sum_{i=1}^{N} a_{ij}\mathbf{u}_{ij}\left(|\mathbf{R}_i - \mathbf{R}_j| - |\mathbf{R}_i^0 - \mathbf{R}_j^0|\right), \tag{10}$$

where a vector $\mathbf{R}_i \equiv \mathbf{R}_i(t)$ specifies a spatial position of the i-th node at an instant t , $\mathbf{R}_i^0$ is the equilibrium position of the i-th node, and $\mathbf{u}_{ij} = (\mathbf{R}_i - \mathbf{R}_j)/|\mathbf{R}_i - \mathbf{R}_j|$ is the direction of relative displacement of the nodes i and j . For small deviations, $\mathbf{r}_i$, from equilibrium, the equations (10) can be linearized (for details, see [24]):

$$\frac{d\mathbf{r}_i}{dt} = -\sum_j \mathbf{\Lambda}_{ij}\mathbf{r}_j,$$

where the linearization matrix $\mathbf{\Lambda}$ consists of 3×3 blocks, $\mathbf{\Lambda}_{ij}$, that represent the

strain tensors. Therefore, for small perturbations, relaxation to equilibrium is described by the sum of independent (normal) exponential modes,

$$\mathbf{r}(t) = \sum_{k=1}^{3M} r_k(0) \exp\{-\lambda_k t\} \mathbf{e}_k \,,$$

where $\mathbf{r}(t)$ denotes the $3N$-component deviation, and λ_k and $\mathbf{e}_k$, $k = 1, 2, \ldots, 3N$ are the eigenvalues and the eigenvectors of the linearization matrix $\mathbf{\Lambda}$, respectively. A slower normal mode corresponds to a smaller eigenvalue.

The spectrum of eigenvalues is of initial interest for the characterization of molecular machines. Studies of elastic networks of biological molecular machines have revealed two specific properties [24]: i) a large spectral gap separating a few (typically, one or two) slowest (soft) relaxation modes from the rest of spectrum with many fast (rigid) relaxation modes, and ii) a low-dimensional manifold with a large basin of attraction on which a molecular machine performs the action. These properties have a clear physical meaning: The spectral gap indicates the possibility for large-scale slow motion, and the low-dimensional attracting manifold implies that the large-scale motion is reproducible given different perturbations.

3.4. Dynamics of a hierarchically folded sample

Here, we construct an elastic network using a contact map of the designed samples: Representative units are chosen to be the network nodes, and links were placed if the space distance between the nodes did not exceed the cutoff radius equal to three polymer units, $R_{cut} = 3$ [8]. The typical shape of a hierarchically folded sample with characteristics of molecular machines and its elastic network are shown in Figures 5a and 5b, respectively.

The typical relaxation spectra of the sample above are shown in Figure 6a. Note a large spectral gap separating the slowest mode, λ_0 , in the spectrum ($\lambda_1/\lambda_0 \approx 7$, $\lambda_2/\lambda_1 \approx 1$). The three-dimensional representation of the dynamic trajectories obtained from numerical solutions of the complete set of non-linear equations (10) is shown in Figure 6b. The representation is constructed with respect to particular three nodes, two of which maximally reflect the slowest relaxation modes (for details, see [24]).

One can see that the dynamics of a hierarchically folded chain indeed have characteristics of molecular machines, despite the fact that the designed samples do not have the structural motives such as α-helixes or β-sheets typical of proteins. The hierarchically folded sample also quickly relaxes to a one-dimensional manifold with a large attracting basin and then slowly moves on the manifold along a well-defined path. It should be noted that the specific dynamics of molecular machines are inherent for significantly anisotropic folds. Indeed, in the trajectories shown in Figure 6, the fast degree of freedom ($\delta u_{23}/u_{23}^{(0)}$) is rigid for approximately two more orders of magnitude than the slowest degrees of freedom ($\delta u_{12}/u_{12}^{(0)}$ and $\delta u_{13}/u_{13}^{(0)}$).

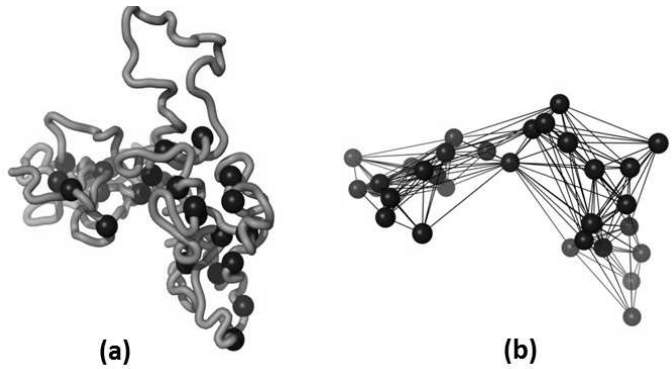

FIGURE 5: Typical sample with characteristics of the molecular machines obtained by the collapse of a polymer chain under the action of a hierarchical potential. (a) The shape of the hierarchically folded structure (red beads represent representative units) and (b) The elastic network of the structure.

3.5. Concluding remarks

In this section, we have considered the overdamped dynamics of an elastic network constructed using a contact map of a crumpled polymer globule formed by relatively short polymer chains. Our results explicitly demonstrate that the dynamics of hierarchically folded globules are similar to those of molecular machines. This observation highlights that crumpled globules may be candidates for designing molecular machines from synthetic polymers. The mechanism that makes it possible to consider the hierarchically folded polymers to be generic molecular machines deals with the sequential energy transfer through a hierarchy of folds from small crumples to the largest crumps. This process produces deformations over scales of the entire globule. The effect is more visible in crumpled globules with an essentially anisotropic shape formed by few large folds.

The virtual structure noted above still does not perform any specific functions such as those of proteins. It is only able to convert energy perturbations of fast degrees of freedom into slow quasi-mechanical motion of large folds. This capability bears a strong resemblance to a heat-engine, and it is possible, in principle, to consider the hierarchically folded chains to be prototypes of handmade molecular machines that can be manipulated by individual molecules. Such an ability definitely provides a new angle to an eternal problem of the origin of life related to overcoming the error threshold in producing and selecting complex molecular structures at prebiotic evolutionary stages (see, for example, [12]; [7]).

These results allow us to put forth a conjecture about the possibility that primary molecular machines can function as a sort of crumpled globule naturally formed under prebiotic conditions. Such molecular machines, even if they are produced from completely different macromolecules from biopolymers, may perform

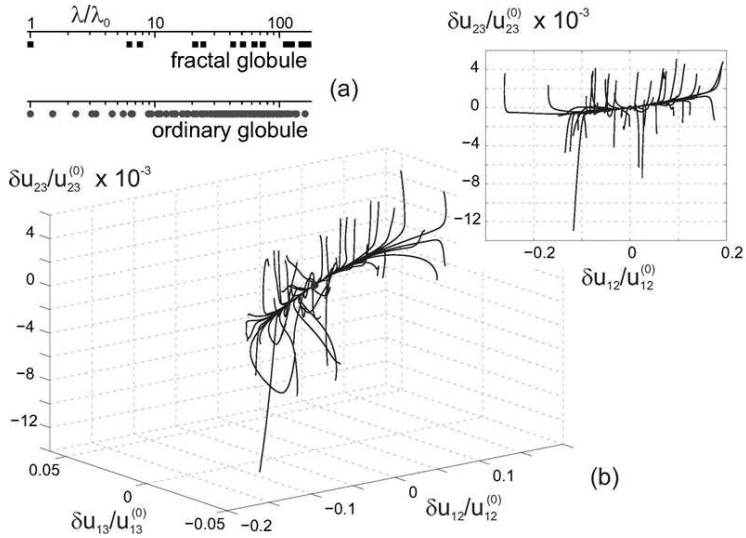

FIGURE 6: (a) The spectrum of relaxation modes of the hierarchically folded globule shown in Figure 5 and an ordinary globule. All of the non-zero eigenvalues are normalized to the smallest eigenvalue λ_0 . Note that there is no spectral gap in the relaxation spectra of an ordinary globule. (b) The dynamic trajectories of the hierarchically folded globule shown in Figure 5a. The axes $\delta u_{12}/u_{12}^{(0)}$ and $\delta u_{13}/u_{13}^{(0)}$ correspond to relative displacements along the two slowest degrees of freedom; the axes $\delta u_{23}/u_{23}^{(0)}$ relate to a fast degree of freedom. The inset shows the slowest mode $(\delta u_{12}/u_{12}^{(0)})$ versus a fast mode $(\delta u_{23}/u_{23}^{(0)})$.

the functions typical of biopolymers. The diversity of primary molecular machines is concerned largely with the number of slow modes manifested in the dimensionality of the attracting manifold on which the action of a molecular machine is performed. Typically, these manifolds are one– or two–dimensional, thereby enabling functional variability without altering the structural archetype. The idea that hierarchically folded globules consisting of relatively short polymer chains may be the primer molecular machines in evolution certainly requires experimental support. However, our results permit to consider the evolutionary scenarios in which the beginning of biological evolution is associated with the appearance of complex functional systems of molecular machines capable of reproduction and autonomous behavior even though the primary molecular machines are taken out of the biological context.

REFERENCES

1. A. ANSARY, J. BERENDZEN, S. F. BOWNE, H. FRAUENFELDER, I, E. IBEN, T. B. SAUKE, E. SHYAMSUNDER, and R. D. YOUNG: *Protein states and proteinquakes*. Proc. Natl. Acad. Sci. USA **82** (1985) 5000-5004.

2. V. A. Avetisov, A. Kh. Bikulov: *Protein ultrametricity and spectral diffusion in deeply frozen proteins*. Physical Reviews and Letters **3** (2008) 387-396.
3. V. A. Avetisov, A. Kh. Bikulov, S. V. Kozyrev: *Application of p-adic analysis to model of breaking of replica symmetry*. J. Phys. A: Math. Gen. **32** (1999) 8785-8791.
4. V. A. Avetisov, A. Kh. Bikulov, S. V. Kozyrev, V. A. Osipov: *p-Adic models of ultrametric diffusion constrained by hierarchical energy landscapes*. J. Phys. A: Math. Gen. **35** (2002) 177-189.
5. V. A. Avetisov, A. Kh. Bikulov, V. A. Osipov: *p-Adic description of characteristic relaxation in complex systems*. J. Phys. A: Math. Gen. **35** (2003) 42394246.
6. V. A. Avetisov, A. Kh. Bikulov, A. P. Zubarev: *Ultrametric random walk and dynamics of protein molecules*. Proc. the Steklov Inst. of Math. **285** (2014) 3-25.
7. V. A. Avetisov, V. I. Goldanskii: *Mirror symmetry breaking at the molecular level*. Proc. Natl. Acad. Sci. USA **93** (1996) 11435-11442.
8. V. A. Avetisov, V. A. Ivanov, D. A. Meshkov, S. K. Nechaev: *Fractal globules: a new approach to artificial molecular machines*. Biophysical Journal **107** (2014) 2361-2368.
9. O. M. Becker, M. Karplus: *The topology of multidimentional protein energy surfaces: theory and applications to peptide structure and kinetics*. J. Chem. Phys. **106** (1997) 1495-1517.
10. A. Kh. Bikulov: *On solution properties of some types of p-adic kinetic equations of form reactiondiffusion*. p-Adic Numbrs, Ultrametric Analysis, and Applications **2** (2981) 187-206.
11. L. A. Blumenfeld: *Problems of biological physics*. Springer-Verlag, Berlin-Heidelberg, 1981.
12. M. Eigen, J. Mc-Cascill, P. Shuster: *Molecular quasi-species*. J. Physical Chemistry **92** (1987) 6892-6899.
13. H. Frauenfelder: *The connection between low-temperature kinetics and life*. In: Protein Structure: Molecular and Electronic Reactivity (R. Austin et.al, eds.), Springer-Veralg, New York 1972, pp. 245–261.
14. A. Yu. Grosberg, A. R. Khokhlov: *Statistical Physics of Macromolecules*. AIP Press, New York 1994.
15. A. Yu. Grosberg, S. K. Nechaev, E. I. Shakhnovich: *The role of topological constraints in the kinetics of collapse of macromolecules*. Journal de Physique **49** (1988) 2095-2100.
16. S. Kaufmann: *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, Oxford, 1993.
17. E. V. Koonin: *The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life*. Biology Direct **2** (2007) 15-31.
18. A. Milchev, W. Paul, K. Binder: *Off-lattice Monte Carlo simulation of dilute and concentrated polymer solutions under theta conditions*. J. Chem. Phys. **99** (1993) 47864798.
19. M. A. Nowak, H. Ohtsuki: *Prevolutionary dynamics and the origin of evolution*. Proc. Natl. Acad. Sci. USA **105** (2008) 14924-14927.
20. G. Parisi: . *Infinite number of order parameters for spin-glasses*. Phys. Rev. Lett. **43** (1979) 1754-1756.
21. D. Raine, P. L. Luisi: *Open questions on the origin of life*. Origin of Life and Evolution of Biosphere **42** (2012) 379-383.
22. R. Rammal, G. Toulose, M. A. Vorasoro: . *Ultrametricity for physicists*. Rev. Mod. Phys. **58** (1986) 765-788.

23. P. J. Steinbach, A. Ansary, J. Berendzen, D. Braunstein, K. Chu, B. R. Cowen, D. Ehrenstein, H. Frauenfelder, J. B. Johnson: *Ligand binding to heme proteins: connection between dynamics and function.* Biochemistry. **30** (1991) 3988-4001.
24. Y. Togashi, A. S. Michailov: *Nonlinear relaxation dynamics in elastic networks and design principles of molecular machines.* Proc. Natl. Acad. Sci. USA. **104** (2007) 8697-8702.
25. D. J. Wales: *Energy landscapes: With applications to clusters, biomolecules and glasses.* Cambridge Univ. Press, Cambridge, 2003.
26. A. Zharikov, S. F. Fischer: *Nonexponential ligand rebinding of CO and O_2 in myoglobin controlled by fluctuations of the protein.* Chem. Phys. Lett. **249** (1996) 459-469.
27. A. Zharikov, S. F. Fischer: *Scaling law for the non-exponential ligand rebinding of CO in myoglobin.* Chem. Phys. Lett. **263** (1996) 749-758.

ULTRAMETRIČNOST PROTEINA I PROJEKTOVANJE SINTETIČKIH MOLEKULARNIH MAŠINA

Ideja o ultrametričnosti proteina predložena pre skoro 35 godina, koja se odnosi na vezivanje CO mioglobinom, je još uvek pod znakom pitanja. U ovom preglednom radu su diskutovani kontradiktorni pokušaji za opisivanje kinetike CO u okviru poznatih pristupa i ultrametričkog opisa. Pregled je proširen novim uglom gledanja na dizajniranje sintetičkih molekularnih mašina koristeći hijerarhiju i samosličnost kao principe dizajniranja.

Ključne reči: *ultrametričnost proteina, molekularne mašine, hijerarhija, samosličnost*