# On a generalized Levinthal's paradox: The role of long- and short range interactions in complex bio-molecular reactions, including protein and DNA folding

Alexey V. Melkikh [a, *], Dirk K.F. Meijer [b]

[a] Ural Federal University, Yekaterinburg, 620002, Mira str. 19, Russia
[b] University of Groningen, The Netherlands

## ABSTRACT

The current protein folding literature is reviewed. Two main approaches to the problem of folding were selected for this review: geometrical and biophysical. The geometrical approach allows the formulation of topological restrictions on folding, that are usually not taken into account in the construction of physical models. In particular, the topological constraints do not allow the known funnel-like energy landscape modeling, although most common methods of resolving the paradox are based on this method. The very paradox is based on the fact that complex molecules must reach their native conformations (complexes that result from reactions) in an exponentially long time, which clearly contradicts the observed experimental data. In this respect we considered the complexity of the reactions between ligands and proteins. On this general basis, the folding-reaction paradox was reformulated and generalized. We conclude that prospects for solving the paradox should be associated with incorporating a topology aspect in biophysical models of protein folding, through the construction of hybrid models. However, such models should explicitly include long-range force fields and local cell biological conditions, such as structured water complexes and photon/phonon/soliton waves, ordered in discrete frequency bands. In this framework, collective and coherent oscillations in, and between, macromolecules are instrumental in inducing intra- and intercellular resonance, serving as an integral guiding network of life communication: the electrome aspect of the cell. Yet, to identify the actual mechanisms underlying the bonds between molecules (atoms), it will be necessary to perform dedicated experiments to more definitely solve the particular time paradox.

© 2017 Elsevier Ltd. All rights reserved.

## Contents

* Corresponding author.
   E-mail addresses: melkikh2008@rambler.ru (A.V. Melkikh), meij6076@planet.nl
(D.K.F. Meijer).

## 1. Introduction

The problem of protein folding is one of the most important problems of molecular biology. A central problem (the so called Levinthal's paradox) is that the protein is first synthesized as a linear molecule that must reach its native conformation in a short time (on the order of seconds or less). The protein can only perform its functions in this (often single) conformation. The problem, however, is that the number of possible conformational states is exponentially large for a long protein molecule. Despite almost 30 years of attempts to resolve this paradox, a solution has not yet been found. A number of authors (see, e.g., Ben-Naim, 2013; Onuchic and Wolynes, 2004; Finkelstein et al., 2017) believe that there is a solution, but they disagree on the reasons. Other scientists (see, e.g., Berger and Leighton, 1998; Davies, 2004) believe that the paradox is not yet resolved.

The issue of folding is typically considered using two fundamentally different approaches that can be called *"biophysical"* and *"geometrical"*. Researchers that use either one of these approaches mostly do not refer to the work based on the other type of approach. The "biophysical" approach uses concepts such as free energy, entropy, and temperature to study protein folding. Simulations of folding are based on statistical physics. The "geometrical" approach does not focus on these parameters; folding is considered geometrically as a part of the broader context of the folding of figures of different topologies. In particular, computational biology has shown that the problem of folding that is based on H-P (hydrophobic-hydrophilic) model belongs to *NP* complexity class (i.e., generally requires an exponential number of steps).

Another major problem, that is essentially ignored in the literature, is the folding and the function of folded DNA because of its much greater length and thus, its much larger number of possible conformational states.

To solve the problem of folding it seems necessary to somehow unite these areas. We must at least discuss the results obtained, using the different approaches and attempt to develop of a single view on the folding problem.

However, the complex interaction of biologically important molecules is connected not only to their folding but also to the possible *reactions* between these molecules. These reactions are in fact the fundamental basis of all the processes that occur in a living system. The scientific field of "molecular docking" can also be regarded as a specialty that is not directly connected to folding processes in life practice. Within this framework, algorithms that can be used to calculate the interaction of the ligand and protein are considered. Complexity also poses a problem to these studies and requires a solution.

In this regard, it seems urgent to develop a more rigorous formulation of the problem of folding and biochemical reactions in general and to discuss the possible solutions in a broader biological context.

## 2. Levinthal's paradox and its possible solutions (biophysical approach)

The process of protein folding is one of the most important problems of molecular biology. According to the first estimates of Levinthal (1968), the average folding time for a long protein molecule is exponentially large because of the large number of conformational degrees of freedom. Levinthal concluded that a random search can, for this reason, not being performed. In that case, what is the folding mechanism? This problem (the so called Levinthal's paradox) has been considered repeatedly (see, e.g., Anfinsen, 1973, Dill, 1985; Shakhnovich and Gutin, 1989, Zwanzig et al., 1992; Berezovsky and Trifonov, 2002 Trifonov and Berezovsky, 2003; Finkelstein and Ptitsyn, 2002; Bai, 2003, 2006; Grosberg, 2002; Grosberg and Khokhlov, 2010; Finkelstein et al., 2017).

Anfinsen (1973) has proposed the hypothesis that the native protein conformation corresponds to attaining the minimum of its Gibbs energy. Yet, from the point of view of thermodynamics and statistical physics, the problem is to understand how such a complex system reaches equilibrium.

However, despite the large number of publications on this issue, researchers disagree not only about the solution to the paradox but also about whether the problem even exists.

In the following we show the calculations underlying the paradox.

First, let us estimate the chain length for which the enumeration problem does not occur, as previously described (Melkikh, 2015). The total number of states of a protein chain can be estimated as (see, e.g., Berezovsky and Trifonov, 2002)

$$3^N.$$

Here, it was assumed that each domain of a protein has 3 different conformations. If we take the maximal possible populations of such molecules as 1050, then we obtain the following:

$$3^N = 10^{50},$$

Thus, $N \approx 10^2$. Moreover, we can consider the fact that each domain in a protein contains several amino acids to obtain the following rough estimate:

$$N \approx 10^3.$$

Longer chains with more information would not be able to find their native conformation through a random search, at least, during the lifetime of the biosphere. If, however $N < < 10^3$, such molecules might find their native conformation by a simple enumeration of variants, and this time must be small (~ 1c) for intracellular processes. As a consequence, $N$ will also be relatively small.

Zwanzig et al., 1992, developed a statistical model of protein

folding as a special case of the general problem of "the first pass". This representation seemed to be more rigorous than previous formulations. Specifically, calculations of the time of the first pass (folding a protein to reach the native configuration) using random walks showed that the Levinthal estimates are correct. However, if a small "decline" of several kT is introduced for locally incorrect configurations, the time for the first pass to reach the fully correct configuration, this time becomes much shorter and will coincide with the experimental time. It was inferred from these calculations that the Levinthal's paradox was in fact solved. Notice however, that such a "decline" just represents a special path of folding.

One possible solution to the Levinthal's paradox is the assumption of the existence of a funnel-like landscape for reaching the lowest state of free energy of the protein (see, for example, Bryngelson and Wolynes, 1987; Onuchic et al., 1997). A relatively smooth landscape, in this modeling, leads to the relatively short time that characterizes the overall protein folding process (see Fig. 1).

The basic idea of this approach is that there is no single path of folding, but that the shape of the energy landscape resembles a funnel with wavy walls (funnel-like) and some traps (Fig. 2). The basic assumption is a minimum of frustration, i.e., a relatively smooth landscape near the native structure. In particular, (Wolynes, 2015) examined the evolutionary aspects of the funnel-like landscape. The possible evolutionary mechanisms of such a landscape were discussed.

Evolutionary aspects of protein folding were also considered by many authors (see, for example, Bloom et al., 2006; Xia and Levitt, 2004, Tiana et al., 2004; Mirny et al., 1998; Gutin et al., 1996).

Some authors argue that there is no paradox at all and that the solution was in fact proposed by Levinthal himself (Rooman et al., 2002; Fernandez et al., 2002).

Some calculations (Berezovsky and Trifonov, 2002) led to the conclusion that the full search of variants of all protein conformations is impossible and that protein folding follows a path that is predetermined by a yet unknown rule of folding that is dependent on the amino acid sequence. The protein folding process was assumed to be connected to the formation of a loop structure of amino acids, with the loops as elementary units of the folding process.

Garbyzinsky et al., 2013 showed that physical theory with biological constrains outlines the "golden triangle", limiting the possible range of folding rates. The golden triangle predicts the maximal size of protein domains that fold under solely thermodynamic control. Larger protein domains fold under kinetic control.

Other investigators (Flory, 1969; Berezovsky et al., 2000; Grosberg and Khokhlov, 2010; Shimada and Yamakawa, 1984; De Gennes, 1990; Berezovsky et al., 2001), also noted the importance of the loop structure of proteins and proposed some scenarios for the evolution of modern proteins from primitive loop-like ancestors that performed relatively simple functions.

Ben-Naim proposed that the solution to the paradox is related to hydrophobic forces (as opposed to hydrophilic) (for details, see Ben-Naim, 2013). Ben-Naim (2012) proposed to divide all the approaches to the folding problem into target-based and cause-based approaches. For example, the author considers the model of Zwanzig et al. (1992) to be one of the most important but also unrealistic attempts. Ben-Naim considered this approach unrealistic because of the similarity to the ideas of cumulative selection of Dawkins (i.e., it implicitly assumes that there is a purpose for folding). According to the author, there is no right or wrong result of folding because evolution has no real goal.

Finkelstein (Finkelstein and Garbuzinsky, 2013) believes that he solved the paradox in his article (Finkelstein and Bardetdinov, 1997) using assumptions about the funnel-like landscape of folding. In this landscape, the folding entropy loss at each step is compensated for by an energy gain. However, this solution avoids astronomical numbers only for the case when the native structure is significantly more stable than nonnative structures. The latter assumption is also important (see also Melkikh, 2015) because the protein must reside in the native conformation for a sufficiently long time; otherwise, the protein will not have proper time to perform its functions.

Understanding the mechanisms of protein folding is important not only for the protein per se, but also for understanding of all higher order structures at stake. For example, Dykeman (2014) noted that the formation of the protein coat is an important step in the life cycle of viruses.

Pandey and Jain (2014), interestingly, believe that protein folding can be considered the second part of the genetic code. Indeed, if the genetic code determines the sequence of amino acids in the protein, then there must be some *rule (code)* that specifies how the sequence folds. Without this second part, the protein cannot perform its functions. The authors divided protein folding models into two classes: *template-based* and *template-free*. The first
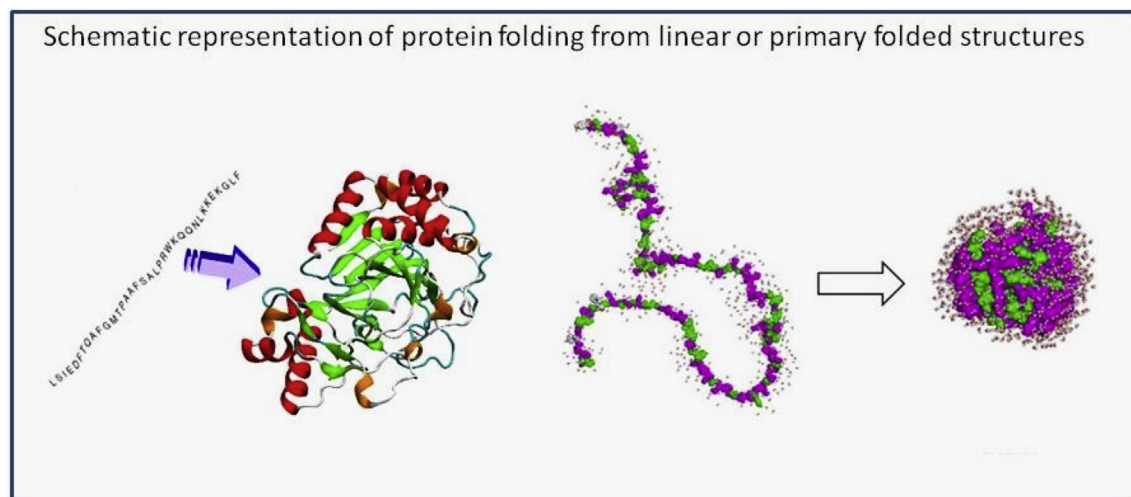


**Fig. 1.** From linear poly-amino-acid chain to folded configuration: possible configurations exceed the billions, yet complete folding in the cell usually is attained within a second (Levinthal's paradox).
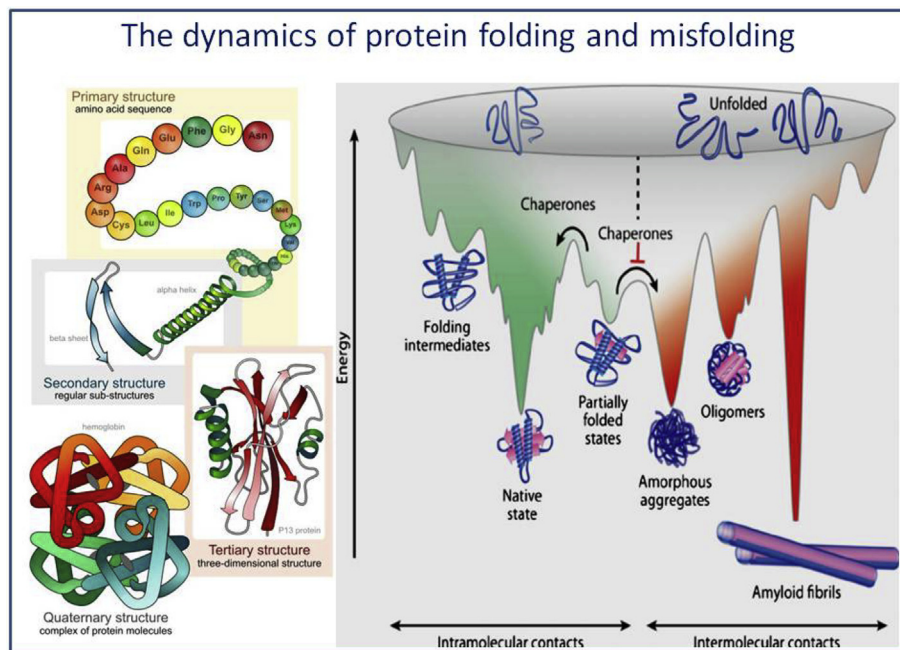
**Fig. 2.** General features of 3-D protein structure (left) and funnel-like landscape showing pathway to minimal energy state for folding and misfolding of proteins (modified from Stefani, 2008).

(homology modeling, threading/fold recognition) requires a known protein structure, but the second does not. The theory of adaptive resonance is based on neural network learning. Note, however, that all these methods are approximate because they do not use the information of specific interactions of individual atoms. As a result, topological problems are largely ignored in these models.

Martinez (2014) notes that the current folding models have weak predictive ability. Natural proteins consist of 100—500 bases, but a random search mechanism cannot provide real-time folding on the order of 1s for these proteins. In the frame of a funnel-like landscape, the author considers a simple kinetic model of protein folding. However, this kinetic model assumed that once a domain reaches the correct state, it stays there. Then, the probability for an amino acid sequence of $N$ residues to be properly folded in a time $t$ can be written as:

$$P = 1 - \left(1 - \frac{1}{2^N}\right)^{t/\tau}.$$

where at every time $\tau$, each residue is perturbed.

*However, it remains unclear how the domain "knows" that it reached the correct state.* After all, an intermediate state is not the minimum energy a state — the minimum is only expected for the whole structure. Note, that this kinetic model also corresponds to cumulative selection, which can only be introduced with certain assumptions (see discussion above). The author believes that if we enter a "decline" to the native conformation, the folding problem can be solved in a realistic time. As already mentioned, such a decline was considered earlier (Zwanzig et al., 1992). It does not follow, however, that such a decline can be introduced due to topological problems.

A number of studies (Brown et al., 2015; Nielsen et al., 2015) discussed various approximate methods for protein folding calculations. In particular, statistical methods, the bat algorithm, swarm intelligence, genetic algorithms and others were considered. The disadvantage of such approximate methods is that they do not tell us anything about how nature solves this problem.

The models of Wolynes, Ben-Naim, Finkelstein and others commonly use the difference in the free energy (free energy gradient) or the Gibbs energy as the driving force for protein folding. It is obvious, however, that the expression for the reactions rates

$$K = \exp\left(-\frac{\Delta G}{kT}\right) \tag{5}$$

uses the concept of "local equilibrium", and this expression is only valid in the case when the system has visited all states, i.e., at large times. However, if proteins typically fold within time scales on the order of seconds and the time for enumerating all conformations of the complex is exponentially large, formula (5) cannot be used (see, also Melkikh, 2015). In this case, it will be the first pass problem, in which the depth of the potential well is not important.

Thus, if it is assumed that the landscape is smoothed at each step, it will be smoothed for the entire molecule also. This smoothing will lead to fast folding and a relatively rapid achievement of equilibrium (which will confirm the validity of the formula). However, if the landscape is not smoothed at each step, there is no reason to apply formula (5) because the characteristic time of folding will be very large, and the system does not have time to visit all of its states. In other words, the protein should follow this equation in the very beginning, although this equation includes the final state.

A number of other issues exist:

1. Whence it follows that the free energy will be compensated at every step?
2. Whence it follows that the native conformation will be more stable than all other conformations?

Thus, all the papers discussed (Finkelstein, Zwanzig and Ben-Naim) (despite some differences) assume a smooth landscape, without which the Levinthal's paradox cannot be resolved.

Numerical modeling of protein folding was also repeatedly

performed (see, e.g., Sali et al., 1994; Lindorff-Larsen et al, 2011). In particular, the folding of a cubic lattice consisted of 27 monomers was simulated (Sali et al., 1994). Such a structure has $10^{16}$ possible conformations, and the search for certain sequences takes $10^7$ steps. The authors believe that a necessary and sufficient feature of fast folding sequences is the presence of an explicit energy minimum. However, the authors do not discuss what potential should be used to achieve such a minimum in nature. The authors also acknowledge that the mechanism must be different for longer proteins.

One paper (Allahverdyan et al., 2013) considered the problem of folding in a different manner - if a machine (device) enumerates the variants, the enumeration process will require energy and will take an exponentially large amount of time. This situation wss considered from the first principles of quantum statistical physics. The result only gives a realistic time of folding for sufficiently short proteins.

In accordance with (Melkikh, 2013, 2014b, 2015), we argue that during folding, topological problems arise that prevent a funnel-like landscape.

Firstly, during the folding of macromolecules energy-equivalent states will always occur; thus, these states will be implemented with equal probability. For example, for a one- or two-component system the presence of such states (folding forks) is quite easy to show (Fig. 3) (see, also Melkikh, 2013).

Due to these equal configurations, the energy landscape is fragmented, and the probability of reaching a single state during folding will be small. In accordance with (Melkikh, 2015), consider a non-periodic chain with n components, each of which has 3 possible conformations. For this case, we can write the probability of incorrect folding in one step that is associated with the presence of the fork in the form:

$$P_E = 1 - \left(1 - \frac{1}{n}\right)^2 .$$

For sufficiently high $n$, this probability has the form:

$$P_E \approx \frac{2}{n} .$$

Then the probability of achieving correct folding for a chain of length $N$ can be estimated using the formula:
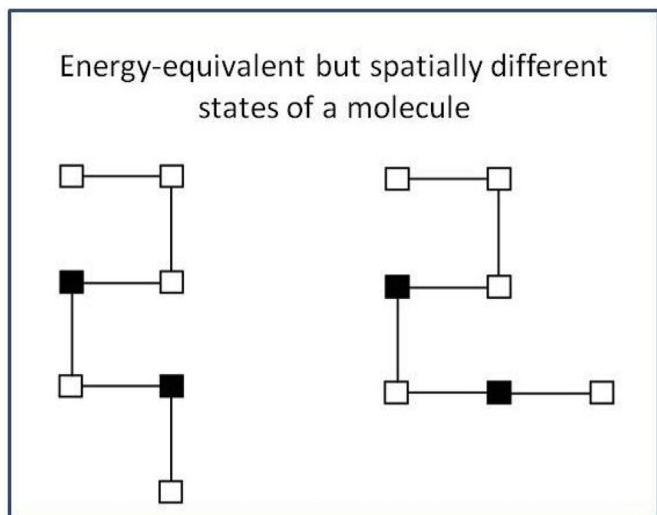
$$P = \left(1 - \frac{1}{n}\right)^{2N} .$$

For example, for $n = 20$ and $N = 100$, we obtain $P = 3.5 \times 10^{-5}$. The dependence $P(N)$ at $n = 20$ is presented in Fig. 4 (Melkikh, 2015).

As shown in Fig. 4, the probability of correct folding appears to be small even for small $N$. In addition, there are restrictions on the value of $n$ (Melkikh, 2015). Indeed, the value of $n$ cannot be too large for the following reason: the total number of different states of energy cannot be greater than $\Delta/kT$ (states, differing by less the kT, will behave the same during folding). Here $\Delta$ is the general energy range of interactions between monomers. The value of $\Delta$ also cannot be large (because these are forces that cause the quaternary structure of proteins), and it is less than 1 eV (40 kT at room temperature).

Thus, due to the presence of degenerate energy (but spatially different) states, a funnel-like landscape is not at stake, and the probability for a molecule to reach its native conformation is low.

In conclusion, we note that many proteins are partially or completely unstructured. Such proteins are preferentially found in higher organisms, though they belong to the oldest RNA-protein complexes: the ribosome and spliceosome. However, unstructured proteins can be dangerous - for example unstructured proteins form the amyloid plaques that are characteristic of Alzheimer's (Stefani, 2004, 2008). Cells control the concentration of disordered proteins; however, the mechanism of such control is unclear because the proposed control assumes that there is a system that somehow distinguishes an ordered protein from disordered. Evidence about the existence of such recognition is inconsistent (see part 4).

Thus, analysis of the current literature clearly indicates that the funnel-like landscape is only a hypothesis. The advantage of this hypothesis is that it is relatively simple, but it should be realized that it contradicts the physical and topological constraints.

Note also that in the vast majority of articles, the authors did not provide specific *interaction potentials*, which are the basis of numerical calculations. Yet, the calculation results will strongly depend on these potentials. Without such specific potentials it will be impossible to test these models.

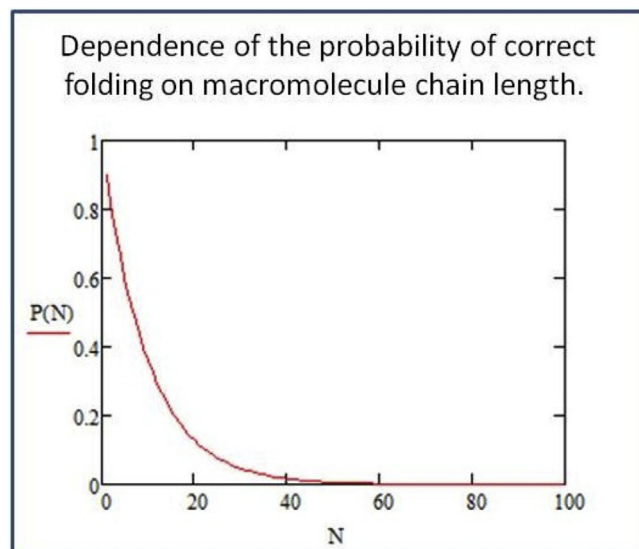The above-mentioned information, about the topological



**Fig. 3.** Energy-equivalent but spatially different states of a molecule.



**Fig. 4.** The dependence of the probability of unmistakable folding on the chain length.

problems of folding, leads to the need for a more detailed examination of this question. This issue will be discussed in section 4. It is clear, however, that the problems arising from protein folding are inherent to a much wider class of molecular intracellular processes, such as DNA and RNA folding, transport of substances, reactions between complex biological molecules, and many others.

### 2.1. Potential long range vibrational mechanisms in protein folding

Preto (2016) recently adequately reviewed the state of art of this field: "Biomolecular systems are expected to exhibit broad vibrational spectral features in the terahertz regime, corresponding to functionally relevant, global, and subglobal collective modes with periods on the picosecond timescale. This hypothesis is supported by the recent emergence of terahertz absorption spectroscopy which has emphasized the broad terahertz absorption of various proteins in dry or partially solvated conditions, as well as in aqueous environments. Terahertz vibrational properties of biomolecules have also been pointed out based on molecular dynamics simulations and normal mode analysis (NMA). Although terahertz spectroscopy has shown a rapid expansion only during the last decade, for example, with the recent accessibility of the so-called "terahertz gap" in a water environment, terahertz vibrational properties of biomolecules were surmised since the 1970s, for example, via Raman spectroscopy experiments. At that time, Fröhlich pointed out on a general theoretical basis that energy provided to terahertz modes of a protein—or possibly of a cell membrane—may not be completely thermalized but redistributed via the nonlinear processes towards specific low-frequency modes of vibration of the biomolecular structure. As reported by many authors, Fröhlich's description relied on a quantum Hamiltonian accounting for the dynamics of phonons of each normal mode of vibration".

This dynamics is assumed to be ruled by linear and nonlinear interactions with a thermal environment (e.g., the cell cytoplasm) as well as linear interactions with an external source of energy, accounting for the available metabolic energy sources (biochemical reactions, ATP hydrolysis, endogenous electromagnetic field).

In these conditions, it can be shown that the system of normal modes will rapidly achieve a steady state characterized by a non-thermal energy distribution, where most of the phonons involved will occupy the modes of lowest frequencies. Applied to biological systems, this effect is equivalent to having the subsystems of a biomolecular structure oscillate cooperatively over long distances at specific frequencies. As emphasized by Reimers et al. (2009), if a coherent effect similar to Fröhlich's effect is experimentally validated, this could lead to numerous applications in physics, biology, and medicine. Indeed, other phenomena that give rise to similar collective properties of a system such as lasing, superconductivity, and Bose-Einstein condensation are of great importance. Since Fröhlich's first proposal, several authors have reported the existence of long-lived excited low-frequency modes in protein structures. Recently, Lundholm et al. (2015) used combined terahertz techniques with a highly sensitive X-ray crystallographic method to visualize the low-frequency vibrational modes in the crystal structure of hen-egg white lysozyme.

Earlier, Pokorny (2004) experimentally observed a strong excitation in the spectrum of vibration of microtubules localized in the 10 MHz range that could also be a further evidence of Fröhlich condensation at lower frequencies. Recently, such interactions were surmised to play an important role in biomolecular organization, as contrary to electrostatic forces, their range could extend far beyond the Debye length given around a few angstroms in the cellular environment. It was, however, found that those interactions are activated only when specific modes of the coupled

system are strongly excited. Nowadays, available experimental devices such as the Fluorescent Correlation spectroscopy have already proved to provide the possibility of detecting long-distance interactions between biomolecules in solution by measuring their diffusion coefficient as a function of their concentration (Nardecchia et al, 2014)".

A soliton is a self-reinforcing solitary wave packet that maintains its shape while it propagates at a constant velocity. Solitons are caused by a cancellation of nonlinear and dispersive effects in the medium.

Goushcha et al. (2014) explained the idea behind the Davydov's soliton as follows: "solitons can be stabilized through the hydrolysis of adenosine triphosphate (ATP) that creates excited vibrational states in the peptide group, allowing vibrational excitation to propagate to the neighboring groups due to dipole interaction between groups. The excitation interacts also with the hydrogen bond, creating a local deformation that is coupled (self-trapped) to the vibrational excitation through a feedback mechanism. If the feedback is strong enough, a new, non-dissipative state associated with vibrational excitation and hydrogen bond distortion will be created and will propagate coherently along the peptide chain.

## 3. Geometrical models of protein folding

Another approach to the problem of folding is associated with geometrical representations this process. Indeed, the folding process can be considered outside the context of the thermodynamic and statistical properties of the molecules; the geometrical representations only consider the internal parameters of the molecules and the distances between them.

The geometrical approach began to develop independently in the 80s and the 90s (see, e.g., Berger and Leighton, 1998). At this time, the influence of the "biophysical" and "geometrical" approaches on one another was quite small. In particular, the idea of the funnel-like landscape was not used in the geometrical approach.

Folding problems in a broader context (outside the context of any molecular structure) have also been considered earlier. One such application of this approach is origami - art of folding paper figures by applying bending operations rather than cutting the paper. In many respects, this type of folding is similar to the actual folding of biologically important molecules; for example, during protein folding, the amino acid sequence remains unchanged, which can be considered to be analogous to the absence of cutting operations. In particular, it was shown in the 1990s (Berger and Leighton, 1998; Crescenzi et al., 1998) that the problem of origami is *NP*-complete (see also Bern and Hayes, 2011).

The absence of such operations can be associated with the problem of topology in the broadest sense — the homeomorphism operation.

We next consider the basic geometrical approaches to the problem of protein folding. One of the most common models of protein folding that is used in geometrical approaches is the hydrophobic-hydrophilic (HP) model. The essence of this model is that all the interactions between protein domains and amino acid residues are reduced to hydrophobic and hydrophilic interactions. The latter leads to the attraction between the domains involved. One of the important goals in the folding problem is to prove its *NP*-completeness (or lack thereof). As noted earlier (Melkikh, 2014b), a two-component model of folding is adequate, because increasing the number of components decreases the differences between the components (i.e., to reduce the difference between the energy levels). In the limit of a large number of components, the energy spectrum becomes almost continuous; thus, components with similar properties cannot be distinguished from each another.

One paper (Berger and Leighton, 1998) showed that the problem of folding under the HP model is *NP*-complete. The similarity between the HP model and the random walk model was noted; in the latter, a point cannot cross its path (SAW - self-avoiding walking). As a result, the problem of protein folding can be reduced to the problem of packing in containers (bin packing problem).

In accordance with (Berger and Leighton, 1998), consider the following formulation of the problem:

There is a polypeptide sequence S of length *N*. Embed the polypeptide in a certain infinite graph G. For a graph G, consider a three-dimensional cubic lattice. Folding S in G is an injective mapping from [1, …, N] in G such that the respective integers correspond to nodes in the whole G. Each node has six neighbors. The energy of the folding of S in G is the number of H-H bonds with opposite sign. The aim is to minimize the energy.

Berger and Leighton (1998) also examined a different formulation of the problem. The goal is to minimize the number of missing H-H bonds. The missed bond is the bond in a pair (u, v), where u contains H, and v contains neither H nor P. The main result of the paper with this formulation was that the problem of determining whether there is a folding within the HP model in the three-dimensional lattice with all 8 missing HH bonds (the absolute minimum) is *NP*-complete. The authors evaluated complexity of this last problem using approximate results.

Note that the situation has not fundamentally changed, that is, if the hypothesis that the native conformation corresponds to the lowest energy minimum is not true. In this case, there is still a search for a particular conformation. It is important that even the degeneracy of the solution (i.e., if the target corresponds to more than one sequence) does not play a fundamental role. It is enough that the power of a set of native conformations was exponentially less than the set power of all conformations. The latter is obviously at stake.

Given a finite sequence S in the alphabet (HP), the integer m and the graph G, the authors (Berger and Leighton, 1998) considered the following question: is there a folding of S in G such that the number of H-H bonds is at least m? If G is $z^3$, then the problem of this folding is *NP*-complete.

The correct HP folding of a string was also considered for an integer n and a finite sequence S on the alphabet (HP) containing $n^3$ number of H. Is there a folding of S in $z^3$ for which all the H correctly fold and become packaged in $n^3$ cube? The proof of the *NP*-completeness of this problem is connected to its reduction to the problem of packing in containers. This problem is defined as follows (Berger and Leighton, 1998):

Consider a limited set of objects U with the size s (u), a positive integer container B, and a positive integer K. The question is whether there is a separation of the set U into separate subsets $U_1$ $\cdots U_K$ such that the sum of the sizes of each subset is equal to or less than B? The modified Bin packing problem is also exists.

As a result, proof of the *NP*-completeness of string folding was obtained (Berger and Leighton, 1998). For the proof, it is shown that a perfect HP string fold is a special case of a more general problem.

In the paper (Crescenzi et al., 1998), it was proved that in the framework of a two-dimensional HP-model folding problem is *NP*-complete. The problem of the NP-completeness can be stated as follows, as described by (Crescenzi et al., 1998):

A two-dimensional grid is a graph$(Z^2, L)$ with a set of nodes $Z^2$ (all the points of the Euclidean plane with integer coordinates) and a set of edges

$$L = \{((x, y), (x', y')) : |x - x'| + |y - y'| = 1\}.$$

Consider a set of strings

$$S = \{s_1, ..., s_m\}$$

from the alphabet 0, 1. The folding of such strings is embedding S into the lattice, i.e., there exists a mapping *f* of set

$$\{(i, j) : 1 \le i \le m, 1 \le j \le |s_i|\}$$

to $Z^2$ such that for all

$$1 \le i \le m, 1 \le j \le |s_i| - 1$$

we have

$$(f(i, j), f(i, j + 1)) \in L.$$

If to fix the folding *f*, than points

$$f(i, j) \text{ and } f(i, j + 1)$$

are considered *f*-neighbors. In folding calculations, a set of edges includes

$$\{(x, y), (x', y')\} \in L$$

such that

a) $(x, x')$ and $(y, y')$ are not *f*-neighbors.
b) Each of these points represents an *f*-mapping of the pair $(i, j)$ such that the j-th symbol of consequence $s_j$ is equal to 1.

The edge of grid $\{(x, y), (x', y')\} \in L$ is said to be loss, if.

A) these points are not f-neighbors
B) exactly one of these two points represents an *f*-mapping of the pair $(i, j)$ such that the j-th symbol of consequence $s_j$ is equal to 1.

In one paper (Crescenzi et al., 1998), the multi-string folding problem was formulated:

Given a set of strings $\{s_1, ..., s_m\} \in 0, 1$ and integer E. Is there a folding with E or fewer losses? If no string does not begin and end with 1, then the total number of folding coincides with a minimum of losses. As a result, the following theorem was proved: The Multi-string folding problem is *NP*-complete. It was shown also, that the problem remains NP-complete even if there is only one string.

Guyeux et al. (2014) noted that the problem of protein folding is insufficiently understood. There are software methods available that can solve the problem for individual proteins, but it is not yet possible for long proteins. These calculations are based on approximate models that minimize the free energy. Note again that in this case, it is crucial which particular approaches are used. What are the potentials of the interaction between atoms, amino acids, and other factors? The paper (Guyeux et al., 2014) considers a 2D lattice model. A related model is the model of random walks without self-intersection - SAW, which is discussed above. It has previously been shown that this problem is *NP*-complete.

The authors note that with the physical method, it is difficult to perform calculations for more than 100 amino acids. Fig. 5 illustrates the folding under the HP model.

In the paper, several SAW tasks, which were characterized by various restrictions that were quite similar to natural restrictions, were considered. The relationships between different SAW tasks were also discussed.

A graph-base approach to the problem of folding was also considered. As a result, the authors concluded that it is impossible to prove the *NP*-completeness of folding for all variants of SAW-
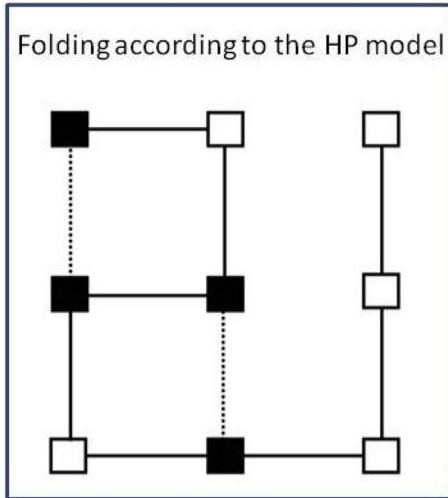
**Fig. 5.** Folding under the HP model.

tasks. However, in the future it should be possible to obtain evidence of *NP*-completeness.

*Moreover, the adequacy of these models is still uncertain because the physical links between atoms (amino acids) are different and cannot be reduced to such simple rules.* For example, the hydration of atoms (ions) could change the results. This limitation can be considered the main disadvantage of the geometrical approach. We postulate that in the future, a hybrid approach based on both geometric and physical models will be possible.

Shaw et al., 2014 considered protein folding on a hexagonal lattice using the HP model. The authors divided the lattice folding models into two classes: simplified lattice models and realistic lattice models. The first class includes the HP-model. Because this model is the most common, many folding algorithms using this framework have been proposed (see, e.g., Hart and Istrail, 1996, 1997; Newman, 2002). The motivation of such a model is the rarity of 90° turns (as in a square lattice) during protein folding.

As a result, the authors proposed two new algorithms for the folding problems within the HP model. In other studies, folding is also considered on a hexagonal lattice (see, for example, Jiang and Zhu, 2005). As shown below, the problem of folding on lattices with different properties is similar to the problem of tiling in many respects.

Note that geometrical approach can be compared to a microscopic description in statistical physics, which deals with a specific topology. To obtain analytic results at some stage, biophysical models neglect these details. In fact, the approach to the folding thus becomes macroscopic. Thus, the Levinthal's paradox, in fact, cannot be solved in the frame of macroscopic description; instead the properties of such interactions between domains were postulated, and these properties can be used to "resolve" the paradox.

Kauffman (2015) showed that the protein secondary structure may be represented as a graph. The presence of knots in the folding can play a critical role. When a molecule with a long chain folds, some locations are connected to each other. Many of these sets exist, so the folding problem is complex. These knots can be designated by bracket symbols. The two chains are isomorphic if they differ only in their choice of the letters, and the chain can be described using the appropriate brackets.

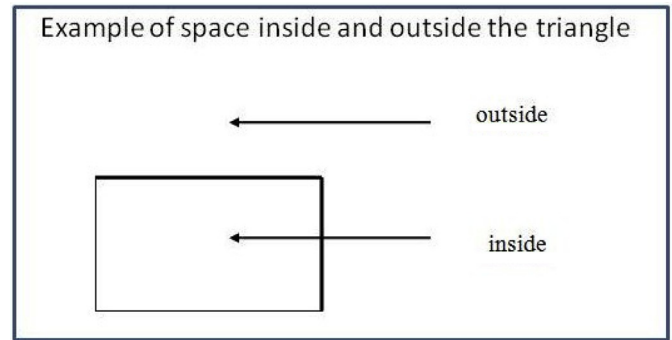An important role Kauffman is assigned to the following symbol:

$\neg$.



**Fig. 6.** An example is the space inside and outside the triangle.

This symbol represents the difference between the outer and the inner region in the broadest sense. An example is the space inside and outside the triangle (Fig. 6):

A general method for the reduction of such symbols was discussed. It has been shown that folding and interaction between the molecules is essentially based on the laws of forms. With respect to the protein folding sequence:

$$C = \langle a | \langle b | \langle c | | c \rangle | b \rangle \langle d | | d \rangle | a \rangle \langle e | | e \rangle$$

will represent pairs of places of the chain link with itself during folding. In this sense, the two chains are considered isomorphic if they differ only in the notation of letters.

According to Kauffman (2015), such a structure can be written in a simpler form:

$$P(C) = < < < < > > < > > < > .$$

The correct notation of brackets for the whole chain can be proved by induction. The concept of "secondary folding" can be introduced on this basis. An example of such a structure, in accordance with (Kauffman, 2015), is shown in Fig. 7.

It should be noted that a similar approach can be applied to any biologically important reaction (see Section 3).

The general conclusions from the biophysical and geometrical models of protein folding can be formulated as follows:
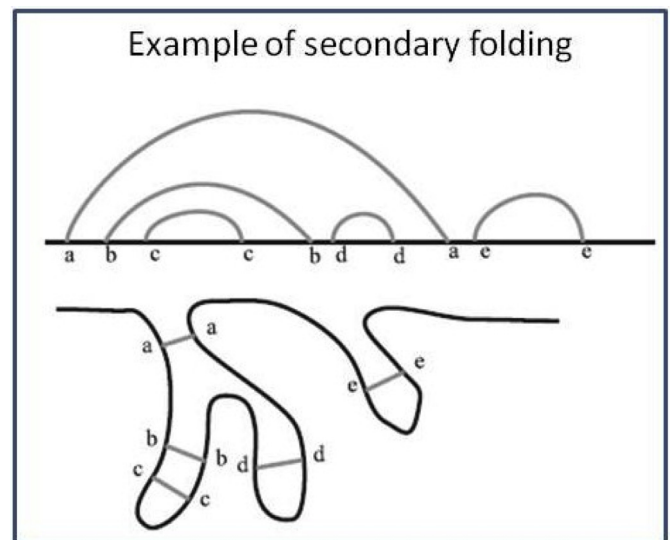


**Fig. 7.** Example of secondary folding (Kauffman, 2015) for the case $\langle a | \langle b | \langle c | | c \rangle | b \rangle \langle d | | d \rangle | a \rangle \langle e | | e \rangle$.

- a funnel-like landscape is not formed and is prevented by a large number of states with identical (degenerate) energy that differ spatially,
- the HP model adequately describes the folding process; conclusions based on this model do not change qualitatively in the case of a larger number of components,
- the NP-completeness of the folding problem can only be proved for special cases, but there is a strong reason to believe that this problem is NP-complete for any potential short-range interaction between atoms.

We also note that in the vast majority of articles on protein folding, the thermodynamics of this process is considered in some detail, but the specific potentials that are used for the calculations are rarely discussed. This omission in itself undermines the credibility of such physical calculations because they inherently depend on these potentials. It is very likely that the authors use some hidden simplifying assumptions, which will simply ignore the topological problems. Thus, the particular solution is for some imaginary and essentially simpler problem rather than for the actual protein folding process.

Note that protein folding is one of the most investigated processes; however, generally speaking, there are many other biochemically important processes. Can we find analogues of the Levinthal's paradox among these processes? What are the possible solutions to this paradox?

## 4. Generalized Levinthal's paradox (reaction-folding paradox) and its connection to other tasks

As noted above, protein folding is not the only complex problem in biological systems. It can be assumed that not only folding but also arbitrary interactions between biologically important molecules pose similar problems.

### 4.1. DNA and RNA folding

The process of protein folding is not the only one that involves complex spatial structures, although such conformations are necessary for protein function. Among these structures, also DNA and RNA can be considered, for which the spatial configuration is known to be very important. We will discuss this question in more detail.

RNA carries out many other essential functions in the cell: RNA regulates gene expression at the transcriptional and translational levels, and this regulation often arises from the structures adopted by various RNA classes, including ribozymes, riboswitches, and RNA—protein complexes (Leamy et al., 2016). According to the authors misfolding and mutations of RNA are characteristics of many cancers and diseases: triplet repeat expansion diseases are associated with Huntington's disease, myotonic dystrophy, and Fragile X syndrome. RNA folding rate depends on cellular environment: small molecule metabolites, polyamines and other species occupy volume and interact with RNAs. Macromolecular crowding can drive the compaction of RNA, while small molecules can either stabilize or destabilize RNAs through interactions with the RNA molecule (Leamy et al., 2016).

DNA is folded into certain nucleoprotein structures in a cell. During the formation of the mitotic chromosome, the DNA of eukaryotic cells is folded several thousand times with great accuracy. Nevertheless, despite the great progress in the study of the DNA folding process, the mechanism of such accurate DNA folding remains unclear.

In eukaryotes, DNA is condensed in chromatin. During intervals between cell divisions chromatin is optimized to provide access to active genes. However, it remains obscure, how exactly such selective access takes place. During division, chromatin is folded in classical chromosomes, where DNA is better structured.

We show that the problem of DNA folding, as well as its function as an information storage, is not trivial and is also even more contradictory and paradoxical than protein folding.

This problem is considered by Melkikh and Khrennikov (2017) from the evolutionary point of view. As we know, DNA in chromosomes represents a condensed medium of high density (see, for example, Teif, Bohinc, 2011), which can be obtained only in the case of a high order of the polymer chains. Let us estimate the number of degrees of freedom of the DNA that is folded in chromosomes. Let the double helix have a length of 3 billion pairs of nucleobases. If we use the persistence length of this polymer (approximately 50 nm) as the domain, the total number of folding variants of such a polymer would all the same be exponentially large. Indeed, if we accept that as a result of bending, a DNA chain can be in at least two different states on the characteristic persistence length, then we obtain for the total number of possible spatial states of DNA:

$$2^{L/L_{pers}}.$$

This is a lower border of the estimation, but the action of enzymes can make this length much smaller. Even for this estimation, the number of states is approximately equal to:

$$2^{2 \times 10^{7}}.$$

This number of states is so astronomically large that it is impossible to enumerate them during the lifetime of the Universe with parallel operation of all living beings that have ever lived on Earth. This means that during folding, DNA has to come to any of the amorphous states, whose number is exponentially large.

Therefore, as exemplified by Melkikh and Khrennikov (2017), each nucleotide is surrounded by approximately six other nucleotides in condensed DNA, but the known interaction potentials between the atoms have a characteristic length of the order of atomic sizes. In this case, the misfolded structure will correspond to the potential well in which the system will stay for a sufficiently long time. Furthermore, there are exponentially many such potential wells (this is shown, for example, for spin glasses). However, the potentials for other molecules (e.g., proteins), which would initially prevent the creation of the wrong spatial structure of DNA, are unknown.

If the mechanism of simple enumerating variants could work during the folding process for relatively short proteins, this method is impossible for DNA because of its significantly larger size.

It is assumed that the control of DNA folding by proteins solves the problem of folding (e.g., histones promote the folding of DNA into a nucleosome); however, the proteins are also molecules for which the same paradox holds (see parts 1 and 2). So, instead of controlling the folding of DNA into regular structures, such proteins could entangle the DNA, because the number of entangled states is much greater than the number of correctly folded states (Melkikh and Khrennikov, 2017).

The problem of DNA and RNA folding was not examined from the physical or geometrical point of view; however, it is obvious that most of the models and their inherent contradictions that were discussed in sections 2 and 3 are fully applicable to these structures. I.e., the DNA and proteins together represent one large macromolecule for which the folding mechanisms are contradictory.

With respect to DNA (RNA) folding, the question of an H-P type model applicability to these structures also arises. As has been shown previously for proteins (Melkikh, 2014b), because the total energy of the interaction is limited and given the requirement that

the characteristic energy of the interaction domains of the protein is significantly large than kT, it turns out that the number of essentially different types of monomers is approximately equal to two. This conclusion is the basis for applying the H-P model to a broad class of systems. In this sense, DNA (RNA) does not represent an exception, because the characteristic energy of their domain interactions is of the same order as for proteins.

### 4.2. Problems of reactions between biologically important molecules including their recognition, copying and transport and potential solutions

The spatial structure of molecules is not only important to the reactions between proteins and DNA, but also for enzymes, transport agents, and similar factors, due to the importance of spatial structure of the molecules for folding. Thus, it can be assumed that the problem of enumerating a large number of variants is also important.

#### 4.2.1. Reactions and recognition

Recognition of molecules during cellular reactions is essential for normal cell function. In particular, it is assumed that the reaction between biologically important molecules obey the principle of "key-lock" (or "hand-glove"). Otherwise, stable and precise work of the cell would be simply impossible due to the large number of "abnormal" reactions. Currently « molecular recognition» and "molecular docking" are special topic in molecular biology and biochemistry (see, e.g., Zsoldos et al., 2007; Mobley and Dill, 2009; Kahraman et al., 2007; Wang and Pang, 2007). Let us consider the basic approaches of this direction.

*4.2.1.1. Molecular docking. Molecular recognition.* These directions are related to each other and represent structure prediction methods; an example is the prediction of the effective interaction between a protein and a ligand. The peculiarity of this area is that the ligand is typically a simpler molecule than the protein. The authors of many studies note that, at present, the accurate prediction of protein-ligand interactions (or more broadly - two biologically important and sufficiently complex molecules) is an unsolved problem. For example, as noted in (Zsoldos et al., 2007), even the number of rotational states for this interaction is on the order of

$$10^{20}.$$

Such a number of possible states is impossible to take into account in calculations regarding two interacting molecular structures. Furthermore, it is not clear how nature chooses from such a number of states. In this sense, this problem is similar to the Levinthal's paradox.

Mobley and Dill (2009) note the current difficulty in creating and studying drugs without performing adequate calculations. In addition, the methods for such calculations are plagued by dilemma of compromising between speed and physical accuracy (as in many other computing tasks in complex systems). Furthermore, the studied object is a biologically important molecule whose properties require improvement.

There are several methods for calculating the optimal binding of a ligand to the protein. The basis of many methods is calculating the free energy of the ligand binding to the complex. It is assumed that the original structures of the protein and the ligand are known. As noted above, the enumeration of all variants of the protein-ligand interaction is not possible; thus, as a rule, only one (or more) location of the ligand binding to the protein is investigated. A protein is considered rigid or only some parts of it are movable.

This type of approach greatly simplifies the task of enumeration because it substantially reduces the total number of investigated conformations.

The calculation algorithm used by the authors (Mobley and Dill, 2009) is based on the following main interactions: hydrogen bonds, hydrophobic interactions, ion pairing and van der Waals forces. As a result, calculations were performed for different conformations of HIV protease. The authors also noted that the role of water is important in calculations because water is inevitably present in all intracellular reactions and because water-ion interactions are quite significant.

To solve the problem of enumerating an exponentially large number of conformations, a priori information about the simulated system seems necessary. This requirement does apply to the problem of molecular docking. Indeed, many of the techniques use databases of proteins and ligands; the information in these databases can accelerate calculation process. For example, in a number of papers (see, for example, Morris, 2006; Kahraman et al., 2007), molecules are modeled using spherical harmonics. Wei et al. (2004) calculated the interaction of thymidylate synthase with a receptor. The Lennard-Jones potential and electrostatic interactions were used.

The paper (Kahraman et al., 2007) noted that the recognition of molecules is mainly based on geometric and electrostatic matching. During evolution, the enzyme and the substrate adjusted to each other – they co-evolved. In protein-ligand interactions, the ligands are usually small molecules with a limited number of conformations. The protein is considered to be partially rigid. Particular attention in calculations is paid to the "pockets" where the ligands bind.

Various programs for calculating ligands binding are discussed in articles (Friesner et al., 2004; Ganesan et al., 2017; Debroise et al., 2017).

Connection of various theories of binding (conformational-induction theory, conformation-selection theory) with the "lock and key" principle were discussed in (Wang and Pang, 2007). The problem of how to achieve the minimum energy for the ligand alone or for the combination of the ligand and the protein was considered. At the heart of many numerical calculations of these configurations is the neglect of small (see, for example, Wang and Pang, 2007) energy gradients.

Morris et al. (2005) note that the concept of "form" has been poorly formalized for molecules. Importantly, these approaches to determine the shape (for example, using Legendre polynomials) lead to faster calculation but inevitably smooth the energy landscape at the molecular level. As noted above with respect to the problem of protein folding, such an operation can actually cause the landscape to become funnel-like; however, this condition causes a very different problem because such smoothing, very likely, does not occur in nature.

One paper (Melkikh, 2015) noted the similarity between the problem of the selectivity of chemical reactions and the problem of protein folding. For example, the problem of reactions between biologically important molecules in the early stages of evolution - during the formation of the first cell - was examined. The formation of protein complexes, generally speaking, can be both harmful and useful for performing work. In the paper (Melkikh, 2015), the interaction between the two sufficiently long molecules was considered, and it was concluded that their interaction kinetics does not differ from folding mechanisms, because the forces acting between monomers of the same and of different molecules have the same nature. Although the repulsion between the individual monomers was also possible, the force of attraction between the monomers should prevail (otherwise, protein folding would be impossible). Then, the complex that results from the two molecules

interacting will be in one of the metastable states. Because the average energy of interaction between the monomers is substantially greater than kT (otherwise, stable folded proteins could not exist), such a metastable state should exist for a sufficiently long time in relation to the characteristic times of the intracellular processes.

Thus, we can consider two interacting (folded) molecules as one (Fig. 8). In this case, their evolution can be considered to be a folding process. As has previously been shown, the funnel-like landscape cannot form during folding; consequently, it is also not formed during the interaction of molecules.

The only difference is that there are topological restrictions during folding related to the fact that long molecules cannot be cut. In the case of a reaction between two molecules, these limitations are weaker, and lead to further increasing the number of possible configurations of the system.

Note that there is a similar problem of antibody recognition in the immune system. This problem is very important for general biological and medical applications. However, antibody recognition mechanisms mostly remain unclear. It is believed that this problem is solved by the so-called "housekeeping genes" and their strong variability. However, mathematical models of the recognition process are absent in the literature, and the problem of the number of variants of antibodies and their possible enumeration in terms of complexity is not discussed.

Thus, during the interaction of two complex biologically important molecules, long-lived intermediate complexes should form; these complexes are not able to perform useful work. The characteristic time to reach the correct "lock and key" ("hand-glove") final state of the reaction should be exponentially large.

### 4.2.2. Transport of molecules and ions

The transport of molecules and ions is also accompanied by processes similar to those of the protein folding. This issue has been considered (Melkikh and Seleznev, 2012; Melkikh, 2013). In particular, it was shown that in the absence of special restrictions,
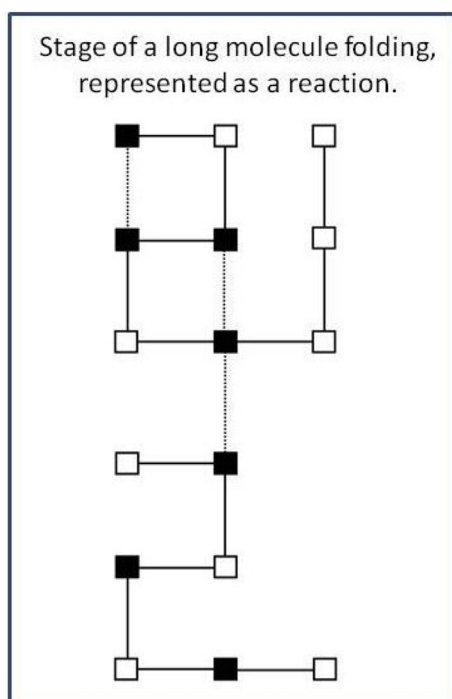


**Fig. 8.** Stage of a long molecule folding, which can be represented as a reaction.

protein (and also RNA) transport is unstable with respect to their interaction with other proteins. For example, the transport of proteins and RNA through the nuclear pore is highly selective; otherwise, the stable organization of such transport would be impossible. That is, in this case, the recognition problems discussed above, will also arise.

As is known, tRNA forms a complex three-dimensional structure, which is unique for each tRNA. If such structures are disturbed, tRNA cannot play role of a transport molecule. The length of tRNA is considerably smaller than that of DNA (73—94 nucleotides); thus, the number of folding variants of such molecules is considerably less than that for DNA. However, the number of all possible reactions of such a molecule with all the molecules in the cell is still exponentially large, i.e., the problem of the selectivity of these reactions remains unresolved for RNA. In the light of all possible reactions, why do only the useful reactions (e.g., tRNA transport through the nuclear pores and biosynthesis of certain amino acids) occur?

Although most of the positive and negative ions in the cell are quite simple (e.g., sodium, potassium, and chlorine), their transport can also be unstable (see, also Melkikh, 2015). For all ions, the transport system requires proteins with sufficiently large mass. In this case, the problem remains unsolved: why do transport proteins transfer a molecule from one side of the membrane to the other with overwhelming probability (with the assistance of ATP, for example) and why are these proteins not involved in any other reactions? Experiments show that the effectiveness of the active transport of ions is very high and approximates 90%. Some studies (Melkikh and Seleznev, 2005; 2006, 2007) developed a model of ion transport that is in good agreement with experiments in terms of the resting potential and intracellular concentration of ions. However, these models are based on the assumption that all other reactions of the transport protein are prohibited. If not, the transport efficiency will be orders of magnitude less and will not correspond to experiments.

Microtubules and microfilaments are also a part of the transport system of the cell. For example, one paper (Pavin, Tolic-Norrelykke, 2014) deals with the problem of how microtubules find their goals in the cell. The authors note that the microtubules move accidentally and directly. Directional movement is carried out with the help of protein complexes (kinetochores). However, this approach, in fact, does not give us a solution because the stable operation of protein complexes themselves is unclear.

Cells stably operate despite the presence of large amounts of protein, RNA and DNA, which can interact with each other. However, the mechanism of this stability remains unclear.

### 4.2.3. Folding, reactions and biological evolution

The problem of the accuracy and selectivity of reactions between biologically important molecules is directly related to the mechanisms of evolution. Indeed, genome regulation, its editing, the work of spliceosomes and other processes directly control the evolution of species.

In one paper (Melkikh, 2015), it was suggested that the same mechanism that is responsible for the exact folding of DNA (as well as all other operations with this molecule, such as transcription, regulation of gene activity, and other processes) could be responsible and for the directed evolution of organisms. In this case, DNA can be regarded as a condensed medium, in which long-range interactions play an important role; additionally, higher order structures (cell, tissue) can be considered in the same way.

The evolution of species is largely determined by the composition and the structure of DNA; thus, the folding mechanism of the molecules, as well as the interaction between the DNA, RNA and proteins, is critical. In particular, one of the most important issues is

the mechanism of gene exchange between two strands of DNA via a sexual process. If the characteristic energy in the exchange of genes is small (of the order of kT), then it will be impossible to ensure accurate copies and gene exchange would occur randomly; thus will occur for parts of the genes that are not relevant. That is, the process of creating a new DNA will be completely broken, resulting in all the expected consequences. However, if the characteristic energy of gene exchange is large (» kT), then the process will not contain randomness. However, as is known, accidents occur during gene exchange. The physical basis of the gene exchange machinery is unclear. What is its efficiency?

On the other hand we can draw an analogy between the problems of folding (reactions) and evolution (see, Table 1) (see also, Melkikh, 2014a; Melkikh and Khrennikov, 2016).

Generally speaking, the very problem of folding or reactions between biologically important molecules could be partially solved in the process of evolution. Indeed, it is known that some proteins are folded hierarchically, i. e. separate parts of the protein chain are folded separately (see, for example, Diaz-Santin et al., 2017). This could, in principle, lead to a reduction in the folding time. However, first, even if such a hierarchical folding is possible, enumeration of variants of the genome for its achievement represents an *NP*-problem (that is, it requires an exponential number of steps). The solution of this problem, as was shown earlier (Melkikh, 2014a; Melkikh and Khrennikov, 2016, 2017), can be obtained in a relatively short time only if there is a priori information about the future states of the system. In this case evolution becomes partially directed. If there are two proteins that, in the course of their undirected evolution, combine into a single complex, the folding of such a complex cannot simply be composed of the folding of individual proteins that existed earlier. In the process of folding, parts of a single molecule interact with each other. As a result, the energy landscape is changing, and the folding problem must be solved for this single molecule separately.

Secondly, the very organization of hierarchical folding will in any case be accompanied by the presence of forks (see Section 1, Fig. 3), which are unavoidable in the presence of only short-range potentials of interaction between atoms of biologically important molecules. The presence of such forks again leads to the necessity of enumeration of an exponentially large number of variants.

One pertinent question is how folding processes of protein (amino-acid chains) did influence biological evolution of (proto) life forms in the creation of first life. This brings up the crucial issue: did the typical neo-darwinian concept of random mutation of DNA, followed by natural selection really played a primary role in this evolution phase? In this classic theory the usual order of events follow the translation rule of DNA-RNA-Proteins. However recent studies challenged this general dogma, proposing the so called "extended evolution synthesis" (Laland et al., 2014, 2017, Miller, 2016; De Loof, 2017; Walker and Davies, 2017). It was envisioned that even primitive cells require some sort of "problem solving" capacity to survive. This by collecting and faithfully storing life-sustaining information from the environment (Walker and Davies, 2017). This process includes not only signal transduction across the plasma membrane, but also stable and retrievable storage of such information.

The latter processes can be readily seen as a direct translation of information in the form of 3-D conformational changes of proteins and in this manner modifying their functional state. It was earlier realized that the synthesis of RNA/DNA may require (precursors of) DNA polymerase and thus protein function. According to the extended theory, it should be considered that a reversed type of translation, in the sense of Protein-RNA-DNA was manifest in the conversion of non-life to life forms, in which the 3-D protein serves as a template for poly-nucleotide synthesis and DNA itself rather became a semi-stable medium for information storing. In other words, another type of natural selection occurred by directed protein modulation, preceding random DNA mutation as the impulse to the subsequent classical type of natural selection.

Yet, the present authors do not see the two opposed sequences of translation as mutual exclusive, since both processes likely have played a role in different phases of (pre)-biological evolution. The "reversed translation" scheme may still be clearly represented in the well-known epigenetic changes in gene expression via protein-guided enzymatic perturbation of DNA and associated proteins such as histones. In fact, epigenetics is the corner stone of the abovementioned extended evolution synthesis theoretical framework. In view of our present proposal, that long distance force fields (photonic/solitonic) may have guided the folding of proteins to their functional modalities, on the basis of an *electromic* constitution of the cell (De Loof, 2017), we envision that such processes may have played a non-trivial role in the creation of first life and further developmental stages in biological evolution.

### 4.2.4. Requirements for potentials

In order for a funnel-like landscape to exist for the interaction of biologically important molecules and their folding, the potential of interaction between the molecules (atoms) must be presented and must satisfy specific additional requirements (Melkikh, 2015):

- first, a potential of interaction should include a sufficient number of distant neighbors (at least an order of magnitude more distant neighbors than nearest neighbors). That is, the characteristic length of the potential should be approximately the size of biologically important molecules. This is the only case in which (at least in principle) it is possible to significantly reduce the number of "forks" (and therefore the probability of error),
- second, this potential should occur only for certain configurations of atoms because the interaction potentials for a condensed matter of arbitrary nature have been well studied, and no additional forces were found, i.e., it must be a collective potential.

However, the known interaction potentials do not meet these requirements. For example, although the Coulomb potential is

**Table 1**
Similarities and differences between folding (reactions) and evolution.

| Property | Folding (reactions) | Evolution |
|---|---|---|
| The number of variants | The number of possible conformations of proteins (possible reactions between biologically-important molecules) is exponentially large | The number of possible variants of the genome is exponentially large |
| The number of correct variants | The number of correct variants is exponentially less than incorrect ones | The number of correct variants is exponentially less than incorrect ones |
| Directivity | Folding (reactions) is partially directed due to long-range forces | Evolution is partially directed |
| Mathematical description | Fokker-Planck equation | Fokker-Planck equation |

relatively long range, it is not selective. The Lennard-Jones and Morse potentials are short range, i.e., at a distance of approximately the size of two atoms, their values are already small (comparable to kT). Chemical (including hydrogen) bonds are always short range.

This issue is confirmed, for example, by the fact that both the folding problem and the problem of protein-protein interactions for long proteins have not yet been resolved (i.e., there are no calculations from first principles for these problems) (see, e.g., Gruebelle, 2010). The current understanding of the interaction of sufficiently long proteins is largely qualitative). It is clear that this problem is much more complicated than the problem of calculating the interaction between a ligand and a protein due to a much larger number of states of the system.

Thus, the existence of multiple spatial structures of biologically important molecules, as well as the many variants of chemical reactions between them, is one of the most important obstacles for understanding the functioning of the cell. There must be some special mechanism that considerably limits the range of variants in such a system. We cannot speak about any significant effectiveness of molecular machines without this mechanism.

### 4.2.5. Aperiodic structures and the problem of protein folding

Quasicrystals discovered by Shechtman began to be studied almost simultaneously with protein folding around the mid-1980s. Quasi-crystals are aperiodic structures with long-range order (e.g., Penrose tiling). They are prepared artificially, but they also exist in nature.

One paper (Marcia, 2006) considered a wide class of aperiodic structures, for which quasi-crystals are a special case. Quasicrystals have been widely used in recent decades (see, e.g., Steinhardt, 2008; Onoda et al., 1988). Note here that the problem of the growth of quasicrystals is largely similar to the problem of protein folding. Indeed, the tiling of the plane by certain figures (building of an aperiodic structure in a space of arbitrary dimension) begins with a particular domain; other domains then join that first domain by specified rules. During the subsequent growth of this type of crystal, the structure remains topologically connected.

The similarity between the problems of protein folding and the growth of quasicrystals has been discusssed (Melkikh, 2014b). A quasicrystal is a quasi-periodic structure with a type of symmetry that is forbidden for ordinary crystals, for example, five-fold symmetry. The problem of the proper growth of the quasicrystal is not trivial because it is impossible to use the simple local rules that are specific to ordinary crystals. At first glance, the growth of quasicrystals and protein folding represent two different problems. However, there is an important similarity between these phenomena. The similarities and differences between these phenomena are presented in Table 2 (Melkikh, 2014b).

Thus, the main similarity between the problems is that, the correct quasicrystal cannot be built and the native conformation of the protein cannot be obtained based on simple local rules that account for the potential of neighboring atoms.

In this sense, the problem of tiling and folding are similar in many respects, and nature solves the task of growing quasicrystals differently than humans solve this problem. This condition does not apply to the traveling salesman problem or the Boolean satisfiability problem – humans invented all these problems. We do not know whether or how nature solves these problems. The fact that in many cases, a number of these problems are *NP*-complete is obvious and does not contradict our intuition (for example, the fact that password hacking requires enumeration of all possible variants in the absence of a priori information about the system).

However, very different approaches are used to address the problem of quasicrystal growth compared to modeling of protein folding. Natural quasicrystals are a special case of a large class of aperiodic structures. Some of these structures are Penrose mosaic and Amman tiles. A Penrose mosaic (1974) represents a tiling of the plane with two different types of rhombus. As a result, symmetry of the fifth order is formed. A Penrose mosaic is the basis for many of quasicrystal growth models. Other examples of aperiodic structures are the Wang domino and Amman tiles. Amman tiles were also used to simulate phase transitions in quasicrystals (see, e.g., Leuzzi and Parisi, 2000; Koch and Radin, 2010; Aristoff and Radin, 2011).

There are a number of quasicrystal models (see, e.g., Onoda et al., 1988; The Toner, 1990; Marcia, 2006; Keys and Glotzer, 2007, Steinhardt, 2008). These models are based on different assumptions about the mechanisms of interaction of elementary structures, which form a quasicrystal. For example, Onoda et al. (1988) proposed a local mechanism in which the tiles are attached to the original cluster. As a result, the correct quasicrystal is formed. It is important that an accidental connection cannot lead to the formation of the correct structure. More precisely, the probability of the formation of such a structure in the event of an accidental connection is exponentially low. Note that although the proposed rules are called "local", they actually include the interaction of rhombi that are in the neighborhood of a given rhombus. However, the rhombus is not an atom, but a certain set of atoms. Thus, the proposed mechanism cannot really be considered "local" because the characteristic sizes of the structures considered in these rule, are much larger than the size of atoms.

A common drawback of the proposed models is that realistic interaction potentials between the atoms that make up the quasicrystal are not considered. In this case, the models of quasicrystal growth are phenomenological and cannot explain the experimental data. Note that this is the same problem as in the case of protein folding!

A paper of one of us (Melkikh, 2014b) noted that long-range interaction is necessary for folding. Fig. 9 shows a part of Penrose tiling, which shows that an identical rhombus is surrounded by different neighbors. However, if there are only short-range forces (extending to a distance of approximately the size of a rhombus or less), the potential is ambiguous because neighboring rhombi are identical and the rhombi that are between them are different (or vice versa).

It remains to be studied what is the potential of interaction between atoms is and if that can provide selective action at a distance.

Another common problem in the growth of quasicrystals and

**Table 2**
Comparative analysis of quasicrystal growth and protein folding.

| Quasicrystal | Protein in the native configuration |
|---|---|
| There is no translational symmetry | There is no translational symmetry |
| Disordered states greatly outnumber quasi-crystalline states. Disordered growth is much more likely. | There exists (a few) unique native configurations. The disordered (entangled) phase is much more likely. |
| Growth of the ideal quasicrystal should take an exponentially large time | Achieving the native configuration must take an exponentially large time |
| Quasiperiodicity takes place | Quasiperiodicity is absent |
| The problem of defects growth exists | The problem of defects growth exists |

## Part of Penrose tiling
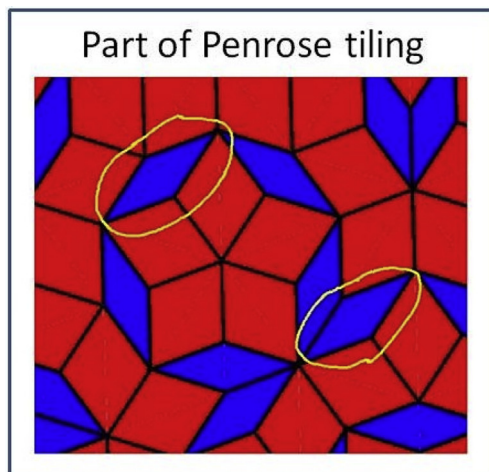
**Fig. 9.** Part of Penrose tiling, where identical rhombus is surrounded by different neighbors.

protein folding, is the behavior of defects. Indeed, if there is a defect in the folding (for example, due to incorrect attachment of one of the rhombus types), the further growth of a quasicrystal cannot stay correct. That is, during the subsequent accretion particles, defects will grow like an avalanche, which will prevent the formation of macroscopic quasicrystals. During protein folding, a similar problem arises - if a defect occurs during the folding process (incorrect position of the domain), such a defect will, with overwhelming probability, make it impossible to continue the correct folding.

Problems of tiling (as in the case of protein folding) could be considered using a geometrical approach, i.e., without the use of physical concepts. The mathematical approaches used in tiling (Frank, 2015; Julien et al., 2015) may be useful for understanding the mechanisms of protein folding and for understanding the mechanisms of a wide class of biochemical reactions.

We also note that the requirements for the potentials of the interaction between atoms for folding and tiling problems are very similar: they both require selective action at a distance.

### 4.2.6. Reaction-folding paradox and related mathematical problems

As has been shown above, the problem of folding is similar to such problems as the bin packing problem and the problem of tiling. We finally consider some related mathematical problems that are similar to the folding problem and show that such a consideration could lead to new formulations of the paradox.

*4.2.6.1. Boolean satisfiability problem.* The *Boolean satisfiability problem* is known to be *NP*-complete (Cook, 1971). For a Boolean formula consisting only of the names of variables, brackets and the operations AND, OR, and NOT, consider the following question: is it possible to assign the value of TRUE or FALSE to all the variables in the formula such that the formula becomes true?

One paper (Kauffman, 2015) proposed to use mathematical logic to describe the protein folding. The variable names will represent the binding sites (domains of proteins, individual atoms), that is, any structure (parts of molecules) that can interact with another. These structures can be numbered in a convenient way. The logical variables are TRUE or FALSE with respect to the correctness or incorrectness of the binding sites. Brackets order the binding, i.e., set the order of the reaction (folding).

The validity of the formula is expressed in the fact that the configuration that results from the reaction between two (or more)

molecules or their folding is native (true). For two interacting macromolecules, a funnel-like landscape is not formed due to presence of forks. This condition inevitably leads to the formation of a large number of long-lived meta-stable states that do not correspond to any effective structures. Imagine that just prior to the interaction, the two molecules had the following characteristic shape (Fig. 10, left column).

After the interaction, specific binding sites and the shape of the resulting molecule will generally change (Fig. 10, right column): As a result of the interaction of two logical structures, a new logic structure will be obtained. Will the truth of each structure individually affect the truth after the reaction? Yet, will this new structure be true?

*4.2.6.2. Topological approaches.* Folding (reaction) is similar, in many respects, also to a wide class of topological problems. The *Poincare hypothesis*, asserts the homeomorphy of every simply connected compact manifold without a boundary to a three-dimensional sphere (Perelman, 2002, 2003a, 2003b). Consider, for example, the conversion of a closed surface to a sphere by a certain topological operations. The deformation of the surface by means of such operations could lead to singularities, that is, to points at which the curvature tends to approach infinity. The deformations occur with the help of the so called Ricci flow.

The similarity with the folding problem is noted - during folding process, similar singularities arise, i.e., situations where a further step-based folding is not possible because the domains of the protein (or other macromolecule) cannot occupy the same place in space (at close distances, repulsive forces are the most significant). Having reached such singularity, and if we want to continue folding, of the particular protein, we will need to make other - collective - operations that change the state of many domains at the same time.

As is well known (Kauffman, 2015 Grosberg and Khokhlov, 2010), sufficiently long biologically important molecules may form knots during folding. The presence of these knots, as well as singularities, leads to increased complexity of the protein folding problem. Thus, for sufficiently long molecules, the existence of collective operations is inevitable. The need for these operations arises from the irregularity of the energy landscape, which was discussed above. In a sense, singularities are special points of the energy landscape in which its curvature exceeds a certain value.

Smoothing a landscape via any "granulation" operation, leads to the fact that its curvature becomes smaller; thus, folding becomes easier. However, such smoothing operation should be prohibited in nature, because the system becomes different when smoothing occurs.

*4.2.6.3. Traveling salesman problem.* We argue that, in principle, the problem of proper folding (reaction) can also be approached via the *traveling salesman problem*. As is well known (see, e.g., Applegate et al., 2007), the traveling salesman problem involves starting from a certain point (the city) and visiting all other cities in the minimum amount of time. It is possible to visit each city only one time. It has been shown (Karp, 1972) that this problem is *NP*-complete. We can suggest the following analogy:

Cities ↔ protein domains, routes between cities ↔ forces acting between the domains, optimal route ↔ native configuration, and movement on route ↔ folding process.

Starting with the initial (unfolded) configuration, the protein gradually passes some states and thus reaches an endpoint. This very path of folding (which remains fixed for folded domains) can be associated with the path on the graph (between cities).

One of the simplest approximate methods for solving the traveling salesman problem is the, so called, greedy algorithm. Its
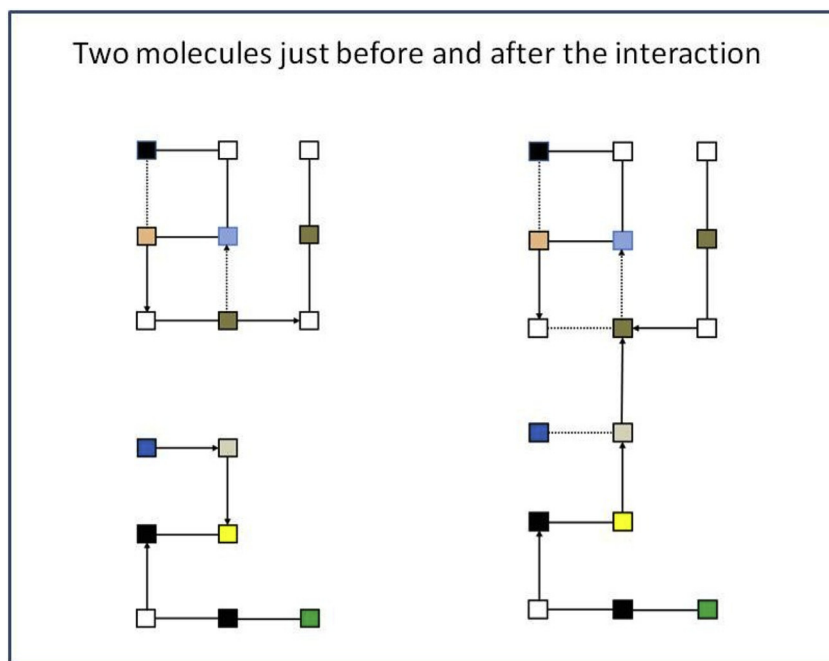
**Fig. 10.** Two molecules just before and just after the interaction.

essence is to go to the nearest town at each step. However, the solution might not exist at all. There are other methods for solving the traveling salesman problem (branch and bound method, method of cutting and other methods). These methods could also be useful for solving the problems of folding and reactions between biologically important molecules.

If we use the greedy algorithm, the solution of this problem, could not obtained at all. In any case, such methods can be considered only an approximate method and does not guarantee a solution. With sufficiently stringent restrictions for errors, this method is not effective.

If the restrictions on the formulation of the problem are weakened (for example, you can visit only a fraction of the cities, etc.), then the problem is greatly simplified. If there is "a priori information" about the solutions of other traversal problems, similar to given problem in the bond topology, we may obtain an approximate solution. Such methods are very similar to the methods for calculating the interaction between a ligand and a protein, as discussed above.

*4.2.6.4. Backpack problem (bin packing problem).* The *backpack problem* was discussed above with respect to a geometrical approach to folding proteins. We show that an arbitrary reaction between complex molecules also has great similarity with this problem.

Consider two interacting molecules with their domains. Both of them together will be a backpack (container). Addition of items in it is a sequential process, which is similar to the reaction between them. The reaction takes place stepwise, when the individual domains interact with each other. Backpack condition at any given moment of time can be characterized by a certain vector. The field of interaction of domains can be characterized by the same vector. After the reaction is over - the backpack is filled. Then we can determine the cost function − the difference of the reactions between the domains from the native reactions. It is possible to apply such a function to the functionality of the resulting complex, which will be the smaller, the more difference from the native

configuration for received molecule.

Note also a possible analogy between the problem of backpack packing and short-range and long-range selective potentials. In the first case all items in the backpack are estimated after every action: putting one item in the backpack. In the second case, the act itself is more complex operation which is putting in the backpack a specific item, if the optimal packaging is known in advance. In many respects, it is similar to the fact that native folding of protein is known in advance and, furthermore, there are certain forces which depend on the states of many domains which tend to bring system to the native configuration.

The precize formulation of the problem in this case is as follows: there are *n* information sequences in an arbitrary initial spatial configuration. It is necessary to find their final configuration which includes both spatial configuration of information sequences themselves and configuration of their complexes for a finite number of steps.

The general conclusion that can be drawn from this section is that folding of biologically important molecules, as well as the reactions between them in many ways are similar to the broad class of known optimization problems. Methods of solving of these problems may therefore be useful for more accurate simulation of biological processes.

## 5. Formulation of a generalized Levinthal's paradox and perspectives for its solution

Based on the collective information stated above, a generalized Levinthal's paradox (reaction-folding paradox) can be formulated. Consider different variants of this paradox (Melkikh, 2014b):

The number of degrees of freedom (primarily conformational) of molecular machines and other molecules in the cell is exponentially large, but in the process of cell functioning, only a small fraction of them are allowed. However, the potentials of the interaction between atoms that are mainly based on interactions with the nearest neighbors do not allow for the selection of a small number of possible states.

This paradox can then be reformulated as follows (Melkikh, 2014b):

On the one hand, the number of different types of particles surrounding a given particle must be large; otherwise, it is impossible to clearly choose a single path. However, the particles are different (primarily due to the interaction energy), and for a given energy range, the typical number of such particles should be small. The implementation of these contradictory requirements using known interaction potentials between particles is impossible.

For a complex system consisting of a large number of atoms, a generalized Levinthal's paradox can be formulated:

Normal cell functioning requires precise biochemical reactions, but the synthesized molecules are so complex that their folding and all possible reactions between them have exponentially many possible variants. In this case, the mechanism for selecting the required reaction from this large number of possible reactions is as yet unclear.

As has been shown above, short-range potentials do not allow a solution to the generalized Levinthal's paradox; thus, we see it as promising to use the following approaches to solve the problem.

## 5.1. Quantum nonlocal effects

It is known that quantum mechanics allows long-range interaction between particles that are connected through their entanglement. Although the required conditions for such interactions are difficult for biological systems, their implementation is, however, possible.

Quantum models of protein folding and interactions of biologically important molecules (Melkikh, 2014b, 2015; Davies et al., 2012; Davies, 2008; Bischof and Del Giudice, 2013, Brizhik et al., 2009; Marchettini et al., 2010; Luo, 2011, 2014) have been proposed.

For example, Davies (2004) proposed a restriction on the time of folding using the Wigner inequality:

$$T < \frac{m_0 a^2 N^3}{\hbar}, \tag{6}$$

where $N$ is the number of amino acids, $m_0$ is the mass of the amino acid, and $a$ is the characteristic size of the amino acid.

Thus, the characteristic time of folding will be proportional to $N^3$. According to Davies, the characteristic size $l$ of a protein is equal to $aN$ only for relatively short proteins. For long proteins, the diameter may be used as a characteristic size of the protein, and the characteristic time may be proportional to $N^{7/3}$. It was found that the experimental and theoretical characteristic times are consistent with each other.

Davis also addressed a number of molecular machines (polymerase, actomyosin, proton pump) and suggested that their work is also determined by the Wigner inequality.

Decoherence is one of the main problems in the discussion of quantum effects in living systems; in most cases, the effects can only be realized in the absence of decoherence. Different authors proposed various methods for controlling decoherence that might be implemented in a living or artificial system (Melkikh, 2015; Davies, 2008), as treated in Meijer and Raggett (2014) and Meijer, 2014. The main methods include the isolation of the molecules by reducing the effective temperature of the molecular complexes, the appearance of subspaces, lack of decoherence due to the Zeno effect, and the correction of errors in quantum computing (see, for example, Knill et al., 1998; Myatt et al., 2000; Igamberdiev, 2004; Kitaev, 2003).

A quantum model of protein folding was proposed in the articles of Luo (2011, 2014). The author notes that the folding mechanism is still not clear; additionally, the temperature dependence of the folding rate is non-Arrhenius. To solve these problems, a quantum model was proposed. In the frame of the model, the author considered the characteristic energies of the chemical bonds in proteins. The author proposed a quantum model of conformational changes of macromolecules, which correspond to quantum transitions between the two states. However, it remained unclear how such transitions help to achieve the native conformation.

A quantum model that takes into account the long-range interactions between atoms has been considered (Melkikh, 2014b). According to this model, the interaction Hamiltonian contains a term that is responsible for the contribution of the distant neighbors of a region surrounding a given particle. It has been suggested that a field $\varphi$ is responsible for this interaction; this field itself can depend on time and system parameters. The model equations can be represented as follows:

The first equation is the Schrödinger equation for a particle; in addition to the usual Hamiltonian, this equation also contained a potential related to the collective interaction of particles.

The second equation describes the dynamics of this type of many-particle potential. This particular potential organizes the collective effects so that the protein folding and other processes discussed above occur with the funnel-like landscape.

Khrennikov and Yurova (2017) proposed to explore the analogy between ontic/epistemic description of quantum phenomena and interrelation between dynamics of conformational and functional states of proteins. They also proposed to apply theory of automata to model the latter dynamics. Protein's behavior is modeled with the aid of two dynamical systems, ontic and epistemic, which describe evolution of conformational and functional states of proteins, respectively. According to the authors this approach does not match to the standard *protein structure-function paradigm*. Space of protein's ontic states (conformational states) is modeled with the aid of p-adic numbers or more general ultrametric spaces encoding the internal hierarchical structure of proteins.

## 5.2. Long-range soliton mediated folding mechanisms

There is a class of systems that is dominated by long-range interaction potentials between particles. Such modalities include systems with gravity, plasma in a magnetic field, system of interacting spins and others (see, e.g., Levin et al., 2014). A feature of such macroscopic systems is that there is no general thermodynamic equilibrium. In the process of evolution, the system enters a quasi-stationary state, which can be quite long. As a result, the techniques used for such systems are very different from the standard approaches in statistical physics. One can assume that during folding, the protein does not approach the equilibrium state or, at least, does not reach it. In this case, the methods of statistical physics, which do not focus on the existence of equilibrium in the system, may be promising for modeling folding (reaction) and for obtaining a solution to the paradox.

The self-organization of the protein folding process augments itself by enhancing the stability of the core against large-scale motions that would unfold the protein, stabilizing the native state through hydrophobic interaction that drives the overall collapse of the chain robustness of the native structure against mutations, etc. Many studies indicated, though, that a unifying concept that combines the basic features governing self-organization of proteins into complex three-dimensional structures in vitro and in vivo is still lacking. This may be attributed to a complexity of the folding process that involves a very large number of interactions (with structural feedback), allowing calculation of only relatively small native structures and applying other critical limitations. Current studies are focused on numerous, as yet, unresolved questions, such as unfolding-refolding reactions and

hidden intermediate role of intra-molecular interactions in folding as well as effects of water dynamics in slaving the chain movements and other topics."

It should be mentioned here that, very likely, proteins do not adopt their native structure in in-vitro conditions, but rather require a complete cellular machinery to correctly fold in vivo in life cells (see Rost, 2003 for many references).

A part of the EM frequencies at stake, were shown by others to be involved in phonon and soliton- (and thus sound-) mediated steering of cellular functions (Pang et al., 2016; Davydov, 1973, 1977; Dotta et al., 2009). It was inferred that the discrete frequency bands (also called *eigenvalues*), as identified in the meta-analysis of the life studies (Geesink and Meijer, 2016a,b), likely reflect a cellular regulation and communication system that may have an evolutionary origin, realizing that due the composition of our planet, EM radiation is a basic property of the planetary environment. Life on earth was formed during billions of years, exposed to, and shaped by the original physical forces such as gravitation, cosmic irradiation, atmospheric electric fields and the terrestrial magnetism. The Schumann resonances at 7.4 Hz and a series of other related frequencies, are an example of oscillation possibly important for life. We conclude that the existing organisms are created to function in harmony with the abovementioned fields and forces which existed when life was born 3 billion years ago.

The migration of soliton energy in molecular systems was first considered by Davydov and Kisluka by the use of a quantum coherent wave theory. Solitons were earlier considered important for energy transfer and storage in biological structures, as described by Davydov (1973) and then by Frohlich, as coherent dipolar propagating waves (see also previous section on potential long range forces). These applications of quantum field theory to biological systems inspired many theoretical physicists to study biological systems with a special interest focused upon tubulin. This filamentous protein is a fundamental building block of the cytoskeleton matter (se Fig. 11).

Microtubules are important components of the cytoskeleton, responsible for cellular organization and information processing. Microtubules of the neurons in the brain might be active components of brain functioning and information processing. Endogenous electromagnetic waves are considered to be moving in the cavity of the microtubules, transporting and carrying information. The relevant mechanism of electromagnetic wave interaction has been suggested to be spontaneous breakdown of symmetry in the biological, well ordered structures. Such interaction occurs with the dipole moments of the molecules in the cell microtubules (see Hameroff and Penrose, 2014; Hameroff, 2016; Sahu et al., 2013, 2014a, 2014b).

These important observations are vividly discussed in relation to the item of consciousness (Meijer and Raggett, 2014; Meijer, 2014; Meijer and Geesink, 2017; Hameroff and Penrose, 2014; Hameroff, 2016). Clearly, they highlight a crucial property of life macromolecules such as proteins, by exhibiting a resonant oscillatory behavior that can obtain a collective character, so that they can communicate in cellular networks through this vibratory state of their 3-D structure (up to thousands of spatial transitions per sec). Such movements can be "frozen" by interactions with hydrophobic agents such as general anesthetics, as reported by Hameroff (2016), for tubulins (see Fig. 11). It should be stressed however that such interaction with anesthetic molecules will occur for all macromolecules in the cell exposing hydrophobic sites such as membrane- and cytoplasmic receptor and channel proteins as well as lipoproteins, histones in cellular organelles. Thus, apart from the known biochemical interactions, proteins can communicate by electromagnetic means (Cosic et al., 2015; De Loof, 2016; Levin, 2012). This, especially in coherent modes that can be perturbed by short- and long range input of energy in the form of photons, phonons and mixed forms such as solitons (Fig. 12).

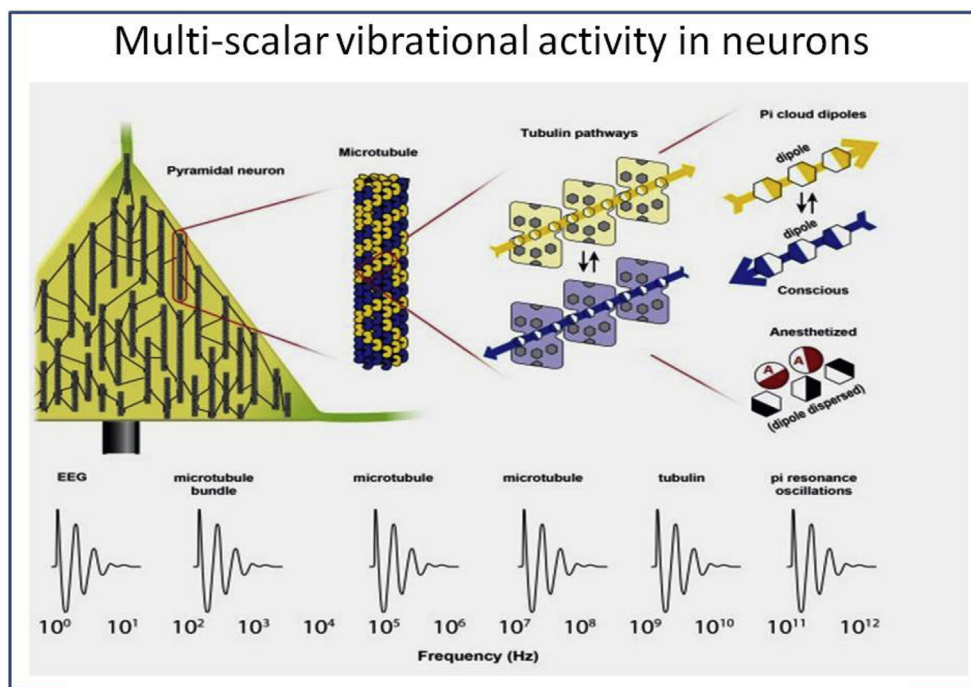Meijer and Geesink (2016a) and Geesink and Meijer (2016b),



**Fig. 11.** Cell (right) and Brain (left) multi-scalar vibrational hierarchy. Top row (left) shows structure (left to right): pyramidal neuron cell body with interior microtubules, a single microtubule, tubulin pathways through pi resonance clouds, and dipole oscillations—resonance transfers and/or spin currents occur, with the upper image showing pi cloud dipole oscillations and the lower image showing anesthetics dispersing dipoles. Bottom row shows dynamics at frequency ranges matching structure in top row. Cell (right) shows intercellular structures containing vibratory macromolecules as an electromic signaling system (modified from Hameroff, 2016).
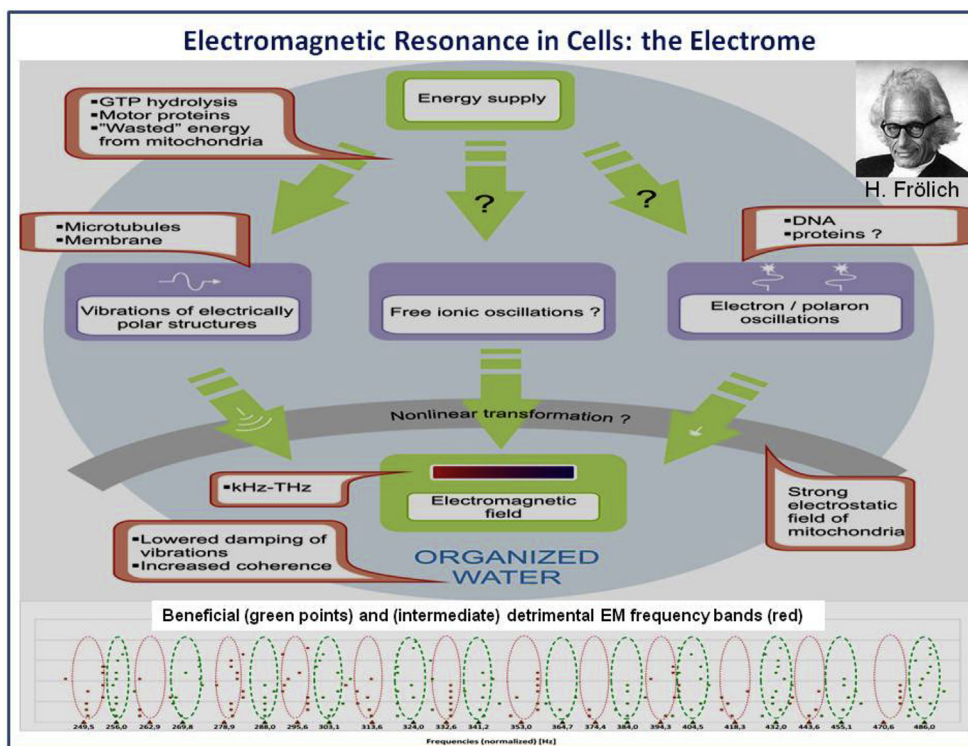
**Fig. 12.** ***Top***: Schematic representation of the ultra-structure of the electromagnetically operating cell, constituting the electrome; sites of oscillatory activity and energy supply are indicated. ***Bottom***: Discrete EM frequency bands derived from 300 biomedical studies in which cells in vivo or in vitro were exposed to electromagnetic fields, plotted on a log frequency scale (Geesink and Meijer, 2016a,b).

gave further experimental support for the Davydov-soliton-states and the an-harmonic and harmonic like frequency zones (Fig. 12, bottom part). The hypothesis that underlying this work on energy transfer in proteins, is that free energy landscape of proteins is multi-funnel-shaped and that protein folding and function involve two steps: a first, kinetic, step in which a specific funnel is selected (most of the times that funnel being the native funnel) and a second step in which the structure relaxes as its free energy is minimized within the funnel selected. As treated above, an algorithm of EM frequencies was earlier proposed (Geesink and Meijer, 2016a, b), that may guide such processes.

Recently, a bio-soliton model was proposed (Geesink and Meijer, 2016b), that predicts which eigenfrequencies of non-thermal electromagnetic waves are life-sustaining and which are, in contrast, detrimental for living cells. The particular effects are exerted by a range of electromagnetic wave eigen-frequencies of one-tenth of a Hertz till Peta Hertz, that show a pattern of twelve bands, and can be positioned on an acoustic frequency scale. The model was substantiated by a meta-analysis of 3000 published papers of biological electromagnetic experiments, in which a spectrum of non-thermal electromagnetic waves were exposed to living cells and intact organisms. Interestingly, the energy released under hydrolysis of ATP molecule, the oscillation energy of the C = O of the peptide groups and the different bending modes of interfacial water molecules fit precisely with the calculated soliton frequencies of the acoustic wave-function, respectively: 0.415 eV, 0.2073 eV and 1660-1693 cm$^{-1}$. Also the 1.1 THz, mentioned by Brizhik (2013), fits exactly with the calculated algorithm of solitons related to a stabilizing coherent frequency.

It can thus be conceived that multiple bio-solitons, of different discrete energies, travel along the primary protein backbone, thereby providing essential long- or short range information, that leads to the final conformational shape of the molecule (see for

folding mechanisms later on). The various solitons could arise from external EM fields, preferably in the polarized form (Geesink and Meijer, 2017a), in relation to a more optimal tissue penetration or being formed internally and organized in coherent domains or collective modes in the cell (see Fig. 12). The question should be raised here how the particular series of traveling solitons translate the fine-tuning forces that finally shape the protein in its functional 3-D configuration.

If special combination of solitonic wave frequencies would play a role, some sort of pre-mordial or evolutionary harmonic information could be responsible. Alternatively, the discrete EM energies could be derived from an addressable memory space for geometric forms of life macromolecules that have been build up in biological evolution and opens the potential to so called backward causation (seen as a strong form of emergence, Meijer, 2012; Meijer and Geesink, 2017). See for the latter aspect also Davies, 2004, 2008, Walker and Davies, 2017 and Grandpierre, 2014. It is of interest that the detection of the abovementioned spectrum of discrete EM frequencies can also be approached by topological geometric modeling in the form of toroidal trajectories (Meijer and Geesink, 2016). This approach is also known from music theory in the so called Tonnetz harmonic analysis.

It should be mentioned here that apart from the coherence promoting solitonic waves, also soliton frequency bands were identified that produce detrimental and de-coherent influences. Thus misfolding or destabilization of proteins (for adequate reviews on this aspect see Stefani, 2004, 2008) could be related to such de-coherent EM radiation fields (Geesink and Meijer, 2017a), as have also been suggested for cancer (Geesink and Meijer, 2017b; Plankar et al., 2011) and even in cataract and Alzheimer disease (Todorova et al., 2016; Stefani, 2004, 2008).

Caspi and Ben-Jacob (2000), suggested that Davidov's solitons, propagating through the backbone of a protein, can mediate

conformational transition and folding of a protein to its native state. A simple toy model indicates that a Non Linear Schrodinger (NLS) field interacts with another field if corresponding to the conformation angles of the protein. The interaction provides the conformation field with the energy needed in order to overcome energy barriers for folding, thus avoiding the need for stochastic thermal activation. Such a transition is therefore a deterministic controlled process.

The solitons involved are compensated for its energy loss by absorption of the energy gained in the folding process. This scenario does not change significantly, even in the presence of imposed disorder, provided that the initial momentum of the soliton is large enough. Transition from a meta-stable to a stable conformational state is usually treated as being thermally induced. Caspi and Ben-Jacob suggested that conformational transition of this sort may rather be mediated by solitons, via a mechanism similar to the one we describe for protein folding.

As mentioned earlier, solitons are seen as localized, non-dispersive excitations, which exist in many nonlinear systems. They are very stable, and therefore can propagate without much energy loss or dispersion to much larger distances than wave-packets of linear waves. The role of solitons in the process of folding (or change of conformation) in proteins would be to provide an efficient mechanism for extracting the energy gained in a single event of local conformational transition and transferring it to a distant location. There, it may be used to activate another transition, and the energy released in that process could be extracted again. This picture is very different from the usual assumption that the energy released in each folding step dissipates to the environment. In this process, large sections of a protein can fold very fast (actually, instantaneously on a time scale of the entire global transition). Moreover, the folding process can be orchestrated deterministically.

The energy that is carried by the soliton, is supposed to be transferred into the conformation field, and this allows it to overcome essential energy barriers and reach the desired ground state. The soliton then receives back some of the energy that was gained in this process, which, at the "steady state", balances its energy loss. This process can go on as long as enough local folding energy is available. Unlike a stochastic process, in which all the energy that is gained at a specific local fold transition dissipates, here a much more efficient mechanism allows a fast and well-controlled conformation transition. The authors suggested that soliton creation is probably induced spontaneously by thermal excitation, or by an external agent such as 'chaperone' enzyme. It might be that solitons are generated in locations of preferable amino-acids sequences. Their propagation might be blocked on different sequences. Therefore, the sequence may not only dictate the final conformation, but also the dynamics of the conformational transition.

Sinkala (2006), also suggested that folding and conformation changes of proteins may be mediated by interaction with solitons which propagate along the molecular chain. In fact, many biological processes in any living organism are associated with conformational changes, which are a result of space propagation of energy and electrons along protein molecules. For example, the energy released (under normal physiological conditions releases 0.42 eV of energy) under hydrolysis of ATP molecule. The major question is what happens to this energy? How does it perform useful work? Is the energy used through non-equilibrium process or does it thermalize and then work through an equilibrium processes? One hypothesis is that in some cases it is transferred along a-helical protein molecules as the vibration oscillation of atoms C=O of peptide groups contained in these molecules.

All of these factors lead to the assumption that energy, released

by ATP, might stay localized and stored in the amide-I vibration, for example see Davydov (1973). He suggested that the amide-I energy could stay localized through nonlinear interactions of the vibrational excitation and the deformation in the protein structure caused by the presence of the excitation. The excitation and the deformation balance each other and form a soliton. Therefore, a soliton is a localized packet of energy.

Protein molecules also transport electrons from donors to acceptors very effectively. This energy can be transferred through oscillations of a physical field that is possibly electromagnetic in its nature. As there is evidence that proteins have certain conducting or semiconducting properties (Davydov, 1973), then charge moving through the protein backbone and passing different energy levels, caused by different amino acid side groups can produce favorable conditions. The general properties of solitons, being solitary waves (waves localized in space) are: 1. they preserves their shape and velocities, 2. they are extremely stable to perturbations (in particular collisions with small amplitude linear waves), 3. they are even stable with respect to collisions with other solitons. In such collusions they can even pass through each other and recover their speed and shape after interaction.

Ilieva et al. (2015) made clear that protein folding is the process of formation of a functional 3-D structure from a random coil: the shape in which amino-acid chains leave the ribosome. Anfinsen's dogma states that the native 3-D shape of a protein is completely determined by protein's amino acid sequence. Despite the progress in understanding the process rate and the success in folding prediction for some small proteins, with presently available physics-based methods it is not yet possible to reliably deduce the shape of a biologically active protein from its amino acid sequence. The protein-folding problem endures as one of the most important unresolved problems in science; it addresses the origin of life itself. Furthermore, a wrong fold is a common cause for a protein to lose its function or even endanger the living organism, as mentioned before.

Soliton solutions of a generalized discrete non-linear Schrödinger equation, obtained from the energy function in terms of bond and torsion angles $\kappa$ and $\tau$, provide a constructive theoretical framework for describing protein folds and folding patterns. Here they studied the dynamics of this process by means of molecular-dynamics simulations. The soliton manifestation is the pattern helix–loop–helix in the secondary structure of the protein, which explains the importance of understanding loop formation in helical proteins.

Krokhotin et al. (2012), finally showed that structural classification indicates that the number of different protein folds is surprisingly small. It also appears that proteins are built in a modular fashion, from a relatively small number of components. It was proposed to identify the modular building blocks of proteins with the dark soliton solution of a generalized discrete nonlinear Schrodinger equation. For this they show that practically all protein loops can be obtained simply by scaling the size and by joining together a number of copies of the soliton, one after another. The soliton has only two loop specific parameters and the authors identified their possible values in Protein Data Bank. They show that with a collection of 200 sets of parameters, each determining a soliton profile that describes a different short loop, they covered over 90% of all proteins with experimental accuracy. They also presented two examples that describe how the loop library can be employed both to model and to analyze the structure of folded proteins.

Protein loops remain a major challenge both in structure classification and prediction. Loops are commonly viewed as apparently random regions with no regular self-similar structure. The authors have shown that loops are not random at all. Their shape is

fully determined and with experimental B-factor accuracy by the dark soliton solution of a generalized discrete nonlinear Schrodinger equation, that has only two loop specific parameters. In particular they have found that the number of different parameter sets, i.e. fundamental loops appears to be no more than 200 and probably is even smaller than 57, if one allows for multiple coverings. When the fundamental loops together with the helices and strands are at ones disposal, the construction of entire folded proteins becomes like a "play with Lego bricks". The authors claim that they can build the entire protein from these modular components by simply putting them together, one after another. Moreover, their quantitative approach is firmly grounded on a physics-based energy function (Table 3).

Cellular plasma water is generally supposed to act as a transfer medium for external electromagnetic waves to bio-molecules. The cellular plasma may exhibit a highly arranged 3-D geometric structure as a liquid crystal, that exhibits surface interactions with macromolecular structures. The absorption spectrum between 1 THz and 10 THz of solvated bio-molecules is sensitive to changes in fast fluctuations of the water network. There is a long range influence on the hydration bond dynamics of the water around binding sites of proteins, and water is shown to assist molecular recognition processes. "Biological water" supports itself by coherent dipolar excitations and terahertz/femtosecond infrared interactions and these dynamics extends well beyond the first solvation shell of water molecules (Chaplin, 2000; Johnson, 2009; Tielrooij et al., 2010). The reorientation of water molecules around ions and interaction with solvated ions slows down during external THz-waves, shifting the absorption peak to typical frequencies and thereby reducing the absorption of radiation at THz frequencies (Geesink and Meijer, 2016b).

## 6. Conclusions

It is shown in the present review that bio-molecular folding and a much wider class of biochemical reactions requires an unrealistically large number of steps to find the correct functional state, irrespective of the various models developed. The reduction of these issues to some classical problems is considered. On this basis, the known folding-reaction paradox was reformulated and generalized by us. The essence of the generalized paradox, that is, in the proposed broader context, lies in the very fact that complex molecules must reach their native configurations (including their related reactions) during an exponentially large time, which clearly contradicts the observed experimental data on folding time.

In our opinion, perspectives for solving the paradox are associated with an explicit incorporation of topology in the physical models of folding. This condition can be achieved, in our view, through the construction of hybrid models based on the geometrical and physical approaches.

We argue that it is crucial in this respect, to approach the item of protein folding from a holistic standpoint, in which all the interacting factors and intrinsic conditions of the intact cell should be taken into account (see Table 3). These include long-range force fields and local cell biological conditions such as structured water complexes, among many other factors. In particular, there is abundant information in literature on the role of soliton waves (phonon/electron quasi particles) that can be instrumental as guiding elements in the various steps in the folding process.

We therefore propose a rather versatile and holistic approach to the problem, based on the embedding of proteins in an integral cellular context, exhibiting, apart from the known genomic and proteomic modalities, a well-defined "electrome" aspect of the cell. In this framework, we consider the protein molecule as being influenced also by various long distance force fields of nature, such

as gravity and electromagnetism, in addition to the intrinsic vibratory states of macromolecules that locally generate coherent excitations in the cell. The recent finding of a set of discrete EM frequency bands Meijer and Geesink, 2016, that either promote or endanger life conditions, could be a key in further studies directed at the morphogenetic aspects of protein folding in a biological evolutionary context.

However, it will be necessary to perform experiments to identify the specific mechanisms underlying the actual forces between molecules (atoms). In particular, the actual problem is the possible existence of a collective action at a distance between the molecules. The presence of such long-range interactions may substantially change the modeling of protein and DNA folding. Moreover, it could lead to a complete revision of our understanding not only of the folding of proteins but also of the work of molecular biological structures in the broadest sense.

### 6.1. Post-submission addendum

Progress in terahertz technology has enabled one to look at biological systems with terahertz radiation, that is, in an energy domain (a few meV), which is of the order of the activation energy of many biological processes. Among these, collective excitations driven by metabolic activity were hypothesized by H. Fröhlich to account for the huge speed of enzyme reactions and for the fast encounters of the cognate partners of biochemical reactions. In fact, according to this hypothesis, collective oscillations of biomolecules, by bringing about giant oscillating dipole moments, would result in resonant (thus selective) electrodynamic forces acting at a long distance.

The particular model assumes that the activation mechanism of collective oscillations can be seen as a Bose-like condensation of the normal vibrational modes of a biomolecule. In this model a biomolecule is considered as an open system through which energy flows under the simultaneous action of an external supply and of dissipation due to radiative, dielectric, and viscous energy losses. Such Bose-like condensation, in the lowest vibrational mode, is predicted to occur when the energy input rate exceeds some threshold value.

We here shortly mention the recent important work of Nardecchia et al. (2017), since it is highlighting the process of far distance oscillation phenomena in proteins.

The authors stipulate that the condensation phenomenon originally proposed by Fröhlich as a quantum process, has been discussed, because the supposed quantum condition can probably be hardly maintained for biomolecules. In fact, the frequency of collective oscillations is expected in the sub THz domain, around $10^{11}$ Hz, so that at room temperature $kT/h\nu \gg 1$, which may rule out a quantum description. For sure electrons on the proteins are still quantized, but since the authors were interested in the global motion of the protein and since the protein studied is quite heavy, while the temperature is relatively high with respect to the frequency ($kT \gg h\nu$), the quantum treatment may be questioned. Of course, this does not apply to molecular orbitals and *light induced electronic transi*tions which are always of quantum nature. Thus an additional, more classical approach on the macromolecular vibratory communication, apart from the solitonic quantum mechanisms proposed by us, should be considered even though experimental evidence of the existence of collective modes of vibration of biomolecules has been provided *at thermal equilibrium* by means of Raman spectroscopy already years ago by Lundholm et al. (2015) this was measured in a quite artificial setting: crystallized proteins. In fact no experimental evidence was hitherto available of the possibility of exciting *out-of-thermal-equilibrium* collective oscillations of bio-molecules in a solvated system.

**Table 3**
A protein molecule in the organism is never alone.

-Proteins bind to transport proteins facilitating nucleus to cytoplasm transport
-Proteins can be associated with cytoplasmic receptor- and chaperone proteins
-Proteins undergo interaction with (hydrated) cytoplasm alkali cations and anions
-Proteins are exposed to various force fields: gravity, dark energy and zero-point energy
-Proteins are semiconducting molecules that can guide photons, phonons and solitons
-Proteins exhibit a 3-D structure that is sensitive to ordered state of interfacial water
-Protein molecules, individually, show specific vibratory/resonance patterns
-Proteins can communicate in functional networks through their resonant states
-Protein morphology can be influenced by external electromagnetic and geo-magnetic radiation
-Protein molecule folding is influenced by coherent vibration domains in the cell
-Protein structure can be modified by quantum entanglement as a long-distant aspect

Nardecchia et al. (2017) unveiled the necessary condition to activate long-range intermolecular electro-dynamic forces as will occur in life conditions, and showed that this requires a constant addition of energy to the system in order to attain such non-equilibrium conditions.

Accordingly, they worked out a *classical version* of the original Frohlich model, finding that, remarkably, also in a classical context a Fröhlich-like phonon condensation phenomenon is possible. This was illustrated by the authors, for an idealized macromolecule, displaying the deviation from energy equipartition among the normal vibrational modes. This was computed by means of a novel classical version of the quantum Fröhlich model, as a function of the input energy amplitude. When this parameter exceeds a threshold value, the lowest frequency mode deviates from equipartition by enhancing its strength at the expenses of the other normal modes. Though still representing a biomolecule in a very idealized way, this model predicts a classical condensation phenomenon, that seems worth to be experimentally investigated.

Of note, the same group earlier demonstrated long range resonant communication of biomolecules, by measuring their diffusion behavior, indicating an elegant experimental method to probe the activation of resonant electrodynamic interactions among biomolecules (Nardecchia et al., 2014).

The present results were partially obtained in the frame of state task of Ministry of Education and Science of Russia 1.4539.2017/8.9.

## References

Allahverdyan, A.E., Hovhannisyan, K.V., Melkikh, A.V., Gevorkian, S.G., 2013. Carnot cycle at finite power: attainability of maximal efficiency. Phys. Rev. Lett. 111, 050601.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. Science 181, 223—230.

Applegate, D.L., Bixby, R.E., Chvatal, V., Cook, W.J., 2007. The Traveling Salesman Problem: a Computational Study. Princeton University Press.

Aristoff, D., Radin, C., 2011. First order phase transition in a model of quasicrystals. J. Phys. A Math. Theor. 44 (25), 255001.

Bai, Y., 2003. Hidden intermediates and Levinthal paradox in the folding of small proteins. Biochem. Biophys. Res. Commun. 305 (4), 785—788.

Bai, Y., 2006. Energy barriers, cooperativity, and hidden intermediates in the folding of small proteins. Biochem. Biophys. Res. Commun. 340 (3), 976—983.

Ben-Naim, A., 2012. Levinthal's question revised, and answered. J. Biomol. Struct. Dyn. 30 (1), 113—124.

Ben-Naim, A., 2013. The Protein Folding Problem and its Solutions. World Scientific Publishing, Singapore.

Berezovsky, I.N., Trifonov, E.N., 2002. Loop fold structure of proteins: resolution of Levinthal's paradox. J. Biomol. Struct. Dyn. 20 (1), 5—6.

Berezovsky, I.N., Grosberg, A.Y., Trifonov, E.N., 2000. Closed loops of nearly standard size: common basic element of protein structure. FEBS Lett. 466, 283—286.

Berezovsky, I.N., Kirzhner, A., Kirzhner, V.M., Trifonov, E.N., 2001. Protein folding: looping from the hydrophobic nuclei. Proteins 45, 346—350.

Berger, B., Leighton, T., 1998. Protein folding in the hydrophobic-hydrophilic (HP) model in NP-complete. J. Comput. Biol. 5 (1), 27—40.

Bern, M., Hayes, B., 2011. Origami embedding of piecewice-linear two-manifolds. Algorithmica 59, 3—15.

Bischof, M., Del Giudice, E., 2013. Communications and the emergence of collective behavior in living organisms: a quantum approach. Mol. Biol. Int. 2013, 1—19. ID 987549.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., Arnold, F.H., 2006. Protein stability promotes evolvability. PNAS 103 (15), 5869—5874.

Brizhik, L., 2013. Biological Effects of Pulsating Magnetic Fields: Role of Solitons arxiv.org/pdf/1411.6576.

Brizhik, L., Del Guidice, E., Jorgensen, S.E., Marchettini, N., Tiezzi, E., 2009. The role of electromagnetic potentials in the evolutionary dynamics of ecosystems. Ecol. Model. 220, 1865—1869.

Brown, M.S., Coker, J.A., Hua, O.M.T., 2015. A Complexity Measurement for de novo Protein Folding. IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. 1—8.

Bryngelson, J.D., Wolynes, P.G., 1987. Spin glasses and the statistical mechanics of protein folding. PNAS 84, 7524—7528.

Cook, S.A., 1971. The complexity of theorem-proving procedure. In: STOC 71 Proceedings of the third annual ACM symposium on Theory of computing, pp. 151—158.

Caspi, S., Ben-Jacob, E., 2000. Conformation changes and folding of proteins mediated by Davidov's soliton. Phys. Lett. A 272 (1—2), 124—129.

Chaplin, M.F., 2000. A proposal for the structuring of water, 24 Biophys. Chem. 83 (3), 211—221.

Cosic, I., Cosic, D., Lazar, K., 2015. Is it possible to predict electromagnetic resonances in proteins, DNA and RNA? Nonlinear Biomed. Phys. 3, 5.

Crescenzi, P., Goldman, D., Papadimitriou, C., Piccolboni, A., Yannakakis, M., 1998. On the complexity of protein folding. J. Comput. Biol. 5 (3), 423—465.

Davies, P., 2004. Does quantum mechanics play a non-trivial role in life? BioSystems 78, 69—79.

Davies, P., 2008. A Quantum origin of life? In: Abbot, D., Davies, P.C.W., Pati, A.K. (Eds.), Quantum Aspects of Life. Imperial College Press, pp. 3—18.

Davies, P., Demetrius, L.A., Tuszynski, J.A., 2012. Implications of quantum metabolism and natural selection for the origin of cancer cells and tumor progression. AIP Adv. 2, 011101.

Davydov, A.S., 1973. The theory of contraction of proteins under their excitation. J. Theor. Biol. 38 (3), 559—569.

Davydov, A.S., 1977. Solitons and energy transfer along protein molecules. J. Theor. Biol. 66 (2), 379—387.

De Gennes, P.G., 1990. Introduction to Polymer Dynamics. Cambridge University Press, Cambridge.

De Loof, A., 2016. The cell's self-generated "electrome": the biophysical essence of the immaterial dimension of Life? Commut. Integr. Biol. 9 (5), e1197446.

De Loof, A., 2017. The evolution of "Life": a metadarwinian integrative approach. Commun. Integr. Biol. 10, e1301335.

Debroise, T., Shakhnovich, E.I., Cheron, N., 2017. A hybrid knowledge-based and empirical scoring function for protein—ligand interaction: SMoG2016. J. Chem. Inf. Model. 57 (3), 584—593.

Diaz-Santin, L.M., Lukoyanova, N., Aciyan, E., Cheung, A.C.M., 2017. Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution. eLife e28384. https://doi.org/10.7554/eLife.28384.

Dill, K., 1985. Theory for the Folding and Stability of Globular Proteins Biochemistry, vol. 24, p. 1501.

Dotta, B.T., Mulligan, B.P., Hunter, M.D., Persinger, M.A., 2009. Evidence of macroscopic quantum entanglement during double quantitative electroencephalographic measurements of friend's vs strangers. NeuroQuantology 7 (4), 548—551.

Dykeman, E.C., 2014. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. PNAS 11 (14), 5361—5366.

Fernandez, A., Belinky, A., Boland, M.M., 2002. Protein folding: where is the paradox? J. Biomol. Str. Dyn. 20, 331—332.

Finkelstein, A.V., Bardetdinov, A.Y., 1997. Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. Fold. Des. 2, 115—121.

Finkelstein, A.V., Garbuzinsky, S.O., 2013. Levinthal's question answered…again? J. Biomol. Struct. Dyn. 31 (9), 1013—1015.

Finkelstein, A.V., Ptitsyn, O.B., 2002. Protein Physics. Academic Press, Oxford.

Finkelstein, A.V., Badretdin, A.J., Galzitskaya, O.V., Ivankov, D.N., Bogatyreva, N.S., Garbuzynskiy, S.O., 2017. There and back again: two views on the protein folding puzzle. Phys. Life Rev. 21, 56—71. http://doi.org/10.1016/j.plrev.2017.01.025.

Flory, P., 1969. Statistical Mechanics of Chain Molecules. Interscience, New York.

Frank, N.P., 2015. Tilings with infinite local complexity. In: Mathematics of Aperiodic Order. Springer, Basel, pp. 223—258.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., Shenkin, P.S., 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47 (7), 1739–1749.

Ganesan, A., Coote, M.L., Barakat, K., 2017. Molecular dynamics-driven drug discovery: leaping forward with confidence. Drug Discov. Today 22 (2), 249–269.

Garbyzinsky, S.O., Ivankov, D.N., Bogatyreva, N.S., Finkelstein, A.V., 2013. Golden triangle for folding rates of globular proteins. PNAS 110 (1), 147–150.

Geesink, J.H., Meijer, D.K.F., 2016a. Quantum wave information of life revealed: an algorithm for EM frequencies that create stability of biological order, with implications for brain function and consciousness. NeuroQuantology 14, 106–125.

Geesink, J.H., Meijer, D.K.F., 2016b. Bio-soliton model that predicts distinct non-thermal electromagnetic radiation frequency bands, that either stabilize or destabilize life conditions. https://arxiv.org/pdf/1610.04855.

Geesink, J.H., Meijer, D.K.F., 2017a. Nature unites first, second and third harmonics to organize coherent electromagnetic frequency patterns that are crucial for health and disease. A Solit. algorithm discrete Freq. ordering Ther. Restor. life Process. 1–9. Available at: ResearchGate, may, 2017. https://www.researchgate.net/profile/Dirk_Meijer4/publication/317003066_Nature_Unites_First_Second_and_Third_Harmonics_to_Organize_Coherent_Electromagnetic_Frequency_Patterns_that_are_Crucial_for_Health_and_Disease_A_Soliton_Algorithm_with_Discrete_Frequencies_for_Orderin/links/591d803ba6fdcc233fce9f7f/Nature-Unites-First-Second-and-Third-Harmonics-to-Organize-Coherent-Electromagnetic-Frequency-Patterns-that-are-Crucial-for-Health-and-Disease-A-Soliton-Algorithm-with-Discrete-Frequencies-for-Orderi.pdf.

Geesink, J.H., Meijer, D.K.F., 2017b. Cancer is promoted by cellular states of electromagnetic decoherence and can be corrected by exposure to coherent non-ionizing electromagnetic fields. A Phys. Model about cell-sustaining cell-decaying Solit. eigen-frequencies 1–45. Available at: Research Gate, April, 2017. https://www.researchgate.net/publication/316058728_Cancer_is_promoted_by_cellular_states_of_electromagnetic_decoherence_and_can_be_corrected_by_exposure_to_coherent_non-ionizing_electromagnetic_fields.

Goushcha, A.O., Hushcha, T.O., Christophorov, L.N., 2014. Self-organisation and coherence in biology and medicine. Open J. Biophys. 4, 119–146.

Grandpierre, A., 2014. Biologocaly organized quantum vacuum and the cosmic origin of cellular life. In: Tymieniecka, Anna-Teresa (Ed.), Phenomenology of Space and Time. Book 1. The Forces of the Cosmos and the Ontopoietic Genesis of Life. Series: Analecta Husserliana, V. 116, pp. 107–133 (Chapter 11).

Grosberg, A., 2002. A few disconnected notes related to Levinthal paradox. J. Biomol. Str. Dyn. 20, 317–321.

Grosberg, A.Y., Khokhlov, A.R., 2010. Giant Molecules: Here, There, and Everywhere, second ed. World Scientific Publishing Company, London.

Gruebelle, M., 2010. Weighing up protein folding. Nature 468, 640–641.

Gutin, A.M., Abkevich, V.I., Shakhnovich, E.I., 1996. Chain length scaling of protein folding time. Phys. Rev. Lett. 77 (No. 27), 5433–5436.

Guyeux, C., Cote, N.M.-L., Bahi, J.M., Bienia, W., 2014. Is protein folding problem really a NP-complete one? First investigaions. J. Bioinform. Comput. Biol. 12, 1350017.

Hameroff, S., 2016. Change the music. Psychotherapy and brain vibrations. Neuro-psychoter 4, 31–35.

Hameroff, S., Penrose, R., 2014. Consciousness in the universe: a review of the 'Orch OR' theory. Mar 11 Phys. Life Rev. (1), 39–78.

Hart, W.E., Istrail, S., 1996. Fast protein folding in the hydrophobic-hydrophilic model within three-eighths of optimal. J. Comput. Biol. 3 (1), 53–96.

Hart, W.E., Istrail, S., 1997. Robust proofs of NP-Hardness for protein folding: general lattices and energy potentials. J. Comput. Biol. 4, 1–22.

Igamberdiev, A.U., 2004. Quantum computation, non-demolition measurements, and reflective control in living systems. Biosystems 77 (1–3), 47–56.

Ilieva, N., Dai, J., Sieradzan, A., Niemi, A., 2015. Solitons and protein folding: an in silico experiment. AIP Conf. Proc. 1684, 030006.

Jiang, M., Zhu, B., 2005. Protein folding on the hexagonal lattice in the HP model. J. Bioinforma. Comput. Biol. 3 (1), 19–34.

Johnson, K., 2009. Water Buckyball Terahertz Vibrations in Physics, Chemistry, Biology, and Cosmology arXiv:0902.2035.

Julien, A., Kellendonk, J., Savinien, J., 2015. On the noncommutative geometry of tiling. In: Mathematics of Aperiodic Order. Springer, Basel, pp. 259–306.

Kahrama, A., Morris, R.J., Laskowski, R.A., Thornton, J.M., 2007. Shape variation in protein binding pockets and their ligands. J. Mol. Biol. 368 (1), 283–301.

Karp, R., 1972. Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.M. (Eds.), Complexity of Computer Computations. Springer, pp. 85–103.

Kauffman, L.H., 2015. Self-reference, biologic and the structure of reproduction. Progr. Biophys. Mol. Biol. 119 (3), 382–409.

Keys, A.S., Glotzer, S.C., 2007. How do quasicrystals grow? Phys. Rev. Lett. 99, 235503.

Khrennikov, A., Yurova, E., 2017. Automaton model of protein: dynamics of conformational and functional states. Prog. Biophysics Mol. Biol. https://doi.org/10.1016/j.pbiomolbio.2017.02.003.

Kitaev, A.Yu, 2003. Fault-tolerant quantum computation by anyons. Ann. Phys. 303, 2–30.

Knill, E., Laflamme, R., Zurek, W.H., 1998. Resilient quantum computation. Science 279, 342–345.

Koch, H., Radin, C., 2010. Modelling quasicrystals at positive temperature. J. Stat. Phys. 138, 465–475.

Krokhotin, A., Niemi, A.J., Peng, X., 2012. Soliton concepts and protein structure. Phys. Rev. E 85, 031906.

Laland, K.N., Wray, G.A., Hoekstra, H.E., 2014. Does evolutionary theory need a rethink? Nature 514, 161–164.

Laland, K.N., Uller, T., Feldman, M.W., Sterelny, K., Müller, G.B., Moszek, A., Jablonka, E., Odling –Smee, J., 2017. The extended evolutionary synthesis: its structure, assumptions and predictions. Proc. R. Soc. B 282, 20151019.

Leamy, K.A., Assmann, S.M., Mathews, D.H., Bevilacqua, B.C., 2016. Bridging the gap between in vitro and in vivo RNA folding. Q. Rev. Biophysics 49, e10.

Leuzzi, L., Parisi, G., 2000. Thermodynamics of a tiling model. J. Phys. A 33, 4215–4225.

Levin, M., 2012. Molecular bioelectricity in developmental biology: new tools and recent discoveries: control of cell behavior and pattern formation by trans-membrane potential gradients. Bioessays 34 (3), 205–217.

Levin, Y., Pakter, R., Rizatto, F.B., Teles, T.N., Benetti, F.C.P., 2014. Nonequilibrium statistical mechanics of systems with long-range interaction. Phys. Rep. 535, 1–60.

Levinthal, C., 1968. Are there pathways of protein folding. J. de Chimie Physique 65, 44–45.

Lindorff-Larsen, K., Piana, S., Dror, R.O., Shaw, D.E., 2011. How fast-folding protein fold. Science 334 (6055), 517–520.

Lundholm, I.V., Rodilla, H., Wahlgren, W.Y., Duelli, A., Bourenkov, G., Vukusic, J., Friedman, R., Stake, J., Schneider, T., Katona, G., 2015. Terahertz radiation induces non-thermal structural changes associated with Fröhlich condensation in a protein crystal. Struct. Dyn. 2, 054702.

Luo, L.F., 2011. Protein folding as a quantum transition between conformational states. Front. Phys. 6 (1), 133–140.

Luo, L.F., 2014. Quantum theory of protein folding. Science China. Phys. Mech. Astron. 57 (3), 458–468.

Marchettini, N., Del Guidice, E., Voeikov, V., Tiezzi, E., 2010. Water: a medium where dissipative structures are produced by a coherent dynamics. J. Theor. Biol. 265, 511–516.

Marcia, E., 2006. The role of aperiodic order in science and technology. Rep. Prog. Phys. 69, 397–441.

Martinez, L., 2014. Introducing the Levinthal's protein folding paradox and its solution. J. Chem. Educ. 91, 1918–1923.

Meijer, D.K.F., 2012. The information universe. On the missing link in concepts on the architecture of reality. Syntropy J. 1, 1–64. http://www.sintropia.it/english/2012-eng-1-1.pdf.

Meijer, D.K.F., 2014. The extended brain: cyclic information flow, in a quantum physical realm. NeuroQuantology 12, 180–200. http://www.neuroquantology.com/index.php/journal/article/view/754/651.

Meijer, D.K.F., Geesink, J.H., 2016. Phonon guided biology. Architecture of life and conscious perception are mediated by toroidal coupling of phonon, photon and electron information fluxes at discrete eigenfrequencies. NeuroQuantology 14 (4), 49–67.

Meijer, D.K.F., Geesink, J.H., 2017. Consciousness in the Universe Is Scale Invariant and Implies the Event Horizon of the Human Brain accepted for publication in NeuroQuantology, sept, 2017.

Meijer, D.K.F., Raggett, S., 2014. Quantum physics in consciousness studies. Quantum Mind Ext. review, 180 pages http://quantum-mind.co.uk/wp-content/uploads/2014/11/Quantum-Ph-rev-def-2.pdf.

Melkikh, A.V., 2013. Biological complexity, quantum coherent states and the problem of efficient transmission of information inside a cell. BioSystems 111, 190–198.

Melkikh, A.V., 2014a. Quantum information and the problem of mechanisms of biological evolution. BioSystems 115, 33–45.

Melkikh, A.V., 2014b. Congenital programs of the behavior and nontrivial quantum effects in the neurons work. Biosystems 119, 10–19.

Melkikh, A.V., 2015. Paradoxes of early stages of evolution of life and biological complexity. Orig. Life Evol. Biospheres 45 (1), 163–171.

Melkikh, A.V., Khrennikov, A., 2016. Quantum-like model of partially directed evolution. Prog. Biophysics Mol. Biol. 125, 36–51.

Melkikh, A.V., Khrennikov, A., 2017. Molecular recognition of the environment and mechanisms of the origin of species in quantum-like modeling of evolution. Prog. Biophysics Mol. Biol. https://doi.org/10.1016/j.pbiomolbio.2017.04.008.

Melkikh, A.V., Seleznev, V.D., 2005. Models of active transport of ions in bio-membranes of various types of cells. J. Theor. Biol. 234 (3), 403–412.

Melkikh, A.V., Seleznev, V.D., 2006. Requirements on models and models of active transport of ions in biomembranes. Bull. Math. Biol. 68 (2), 385–399.

Melkikh, A.V., Seleznev, V.D., 2007. Models of active transport of neurotransmitters in synaptic vesicles. J. Theor. Biol. 248 (2), 350–353.

Melkikh, A.V., Seleznev, V.D., 2012. Mechanisms and models of the active transport of ions and the transformation of energy in intracellular compartments. Prog. Bioph. Mol. Biol. 109 (1–2), 33–57.

Miller Jr., W.B., 2016. Cognition, information fields and hologenomic entanglement: evolution in light and shadow. Biology 5 (2), 21.

Mirny, L.A., Abkevich, V.I., Shakhnovich, E.I., 1998. How evolution makes proteins fold quickly? PNAS 95, 4976–4981.

Mobley, D.L., Dill, K.A., 2009. Binding of small-molecule ligands to proteins: "What you see" is not always "What you get". Structure 17, 489–498.

Morris, R.J., 2006. An evaluation of spherical designs for molecular-like surfaces. J. Mol. Graph. Model 24, 356–361.

Morris, R.J., Najmanovich, R.J., Kahraman, A., Thornton, J.M., 2005. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding

pocket and ligand comparisons. Bioinformatics 21 (10), 2347—2355.

Myatt, C.J., King, B.E., Turchette, Q.A., Sackett, C.A., Kielpinski, D., Itano, W.M., Monroe, C., Wineland, D.J., 2000. Decoherence of quantum superpositions through coupling to engineered reservoirs. Nature 403, 269—273.

Nardecchia, I., Spinelli, L., Preto, J., Gori, M., Floriani, E., Jaeger, S., Ferrier, P., Pettini, M., 2014. Experimental detection of long-distance interactions between biomolecules through their diffusion behavior: numerical study. Phys. Rev. E. 90 (2), 022703.

Nardecchia, I., Torres, J., Lechelon, M., Giliberti, V., Ortolani, M., Nouvel, P., Gori, M., Donato, I., Preto, J., Varani, L., Sturgis, J., Pettini, M., 2017. Out-of-equilibrium collective oscillation as phonon condensation in a model protein. https://arxiv.org/pdf/1705.07975.pdf.

Newman, A., 2002. A new algorithm for protein folding in the hp model. In: Proceedings of the Thirteenth Annual ACM-siam Symposium on Discrete Algorithms, pp. 876—884.

Nielsen, S.S., Danoy, G., Jurkowski, V., Laredo, J.L.H., Schneider, R., Talbi, E.-G., Bouvry, P., 2015. A novel mulvi-objectivisation approach for optimising the protein inverse folding problem. EvoApplications 2015. Lecture Notes in Computer Science, vol. 9028. In: Mora, A., Squillero, G. (Eds.), Applications of Evolutionary Computation. Springer, Cham.

Onoda, G., Steinhardt, P.J., DiVincenzo, D.P., Socolar, J.E.S., 1988. Growing perfect quasicrystals. Phys. Rev. Lett. 60, 2653—2656.

Onuchic, J.N., Wolynes, P.G., 2004. Theory of protein folding. Curr. Opin. Struct. Biol. 14, 70—75.

Onuchic, J.N., Luthey-Schulten, Z., Wolynes, P.G., 1997. Theory of protein folding: the energy landscape perspective. Ann. Rev. Phys. Chem. 48, 545—600.

Pandey, G.S., Jain, R.C., 2014. An *Ab Initio*-ART approach for protein folding problem. ICACCI, 2014 Adv. Comput. Commun. Inf.. https://doi.org/10.1109/ICACCI.2014.6968257.

Pang, X.F., Chen, S., Wang, X., Zhong, L., 2016. Influences of electromagnetic energy on bio-energy transport through protein molecules in living systems and its experimental evidence. Int. J. Mol. Sci. 17, 1130.

Pavin, N., Tolic-Norrelykke, I.M., 2014. Swinging a sword: how microtubules search for their targets. Systems and Synthetic Biology 8 (3), 179—186.

Perelman, G., 2002. The Entropy Formula for the Ricci Flow and its Geometric Applications. *arXiv*:math.DG/0211159.

Perelman, G., 2003a. Ricci Flow with Surgery on Three-manifolds. *arXiv*:math.DG/0303109.

Perelman, G., 2003b. Finite Extinction Time for the Solutions to the Ricci Flow on Certain Three-manifolds arXiv:math.DG/0307245.

Plankar, M., Jerman, I., Krasovec, R., 2011. On the origin of cancer: can we ignore coherence? Prog. Biophysics Mol. Biol. 106, 380—390.

Pokorny, J., 2004. Excitation of vibrations in microtubules in living cells. Bioelectrochemistry 63 (1), 321—326.

Preto, J., 2016. Classical investigation of long range coherence in biological systems. Chaos 26, 123116.

Reimers, J.R., McKemmish, L.K., McKenzie, R.H., Mark, A.E., Hush, N.S., 2009. Weak, strong, and coherent regimes of Fröhlich condensation and their applications to terahertz medicine and quantum consciousness. PNAS 106 (11), 4219—4224.

Rooman, M., Dehouck, Y., Kwasigroch, J.M., Biot, C., Gilis, D., 2002. What is paradoxical about Levinthal paradox? J. Biomol. Str. Dyn. 20, 327—329.

Rost, B., 2003. Neural network predict protein structure: hype or hit? In: Frasconi, Paolo, Shamir, Ron (Eds.), Artificial Intelligence and Heuristic Methods for Bioinformatics. IOS Press, Amsterdam, pp. 34—50.

Sahu, S., Ghosh, S., Ghosh, B., Aswani, K., Hirata, K., Fujita, D., Bandyopadhyay, A.,

2013. Atomic water channel controlling remarkable properties of a single brain microtubule: correlating single protein to its supramolecular assembly. Biosens. Bioelectron. 47, 141—148.

Sahu, S., Ghosh, S., Fujita, D., Bandyopadhyay, A., 2014a. Live visualizations of single isolated tubulin protein self-assembly via tunneling current: effect of electromagnetic pumping during spontaneous growth of microtubule. Sci. Rep. 3—4, 7303.

Sahu, S., Ghosh, S., Fujita, D., Bandyopadhyay, A., 2014b. Live visualizations of single isolated tubulin protein self-assembly via tunneling current: effect of electromagnetic pumping during spontaneous growth of microtubule. Sci. Rep. 4, 7303.

Sali, A., Shakhnovich, E., Karplus, M., 1994. How does a protein fold? Nature 369, 248—251.

Shakhnovich, E.I., Gutin, A.M., 1989. Relaxation to equilibrium in the random energy model. Europhys. Lett. 9 (6), 569—574.

Shaw, D.L., Islam, A.S.M.S., Rahman, M.S., Hasan, M., 2014. Protein folding in HP model on hexagonal lattices with diagonals. BMC Bioinformatics, 15 (Suppl. 2), S7.

Shimada, J., Yamakawa, H., 1984. Ring-closure probabilities for twisted wormlike chains. Application to DNA. Macromolecules 17, 689—698.

Sinkala, Z., 2006. Soliton/exciton transport in proteins. J. Theor. Biol. 241, 919—927.

Stefani, M., 2004. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. BBA 1739, 5—25.

Stefani, M., 2008. Protein folding and misfolding on surfaces. Int. J. Mol. Sci. 9, 2515—2542.

Steinhardt, P.J., 2008. How does your quasicrystal grow? Nature 452 (6), 43—44.

Teif, V.B., Bohinc, K., 2011. Condensed DNA: condensing the concepts. Progr. Biophys. Mol. Biol. 105 (3), 208—222.

Tiana, G., Shakhovich, B.E., Dokholyan, N.V., Shakhnovich, E.I., 2004. Imprint of evolution on protein structures. PNAS 101 (9), 2846—2851.

Tielrooij, K.J., Garcia-Araez, N., Bonn, M., Bakker, H.J., 2010. Cooperativity in ion hydration. Science 328, 1006—1009.

Todorova, N., Bentvelzen, A., English, N.J., Yarovsky, I., 2016. Electromagnetic-field effects on structure and dynamics of amyloidogenic peptides. J. Chem. Phys. 144, 085101.

Toner, J., 1990. Growth and kinetic roughening of quasicrystals and other incommensurate systems. Phys. Rev. Lett. 64 (8), 930.

Trifonov, E.N., Berezovsky, I.N., 2003. Evolutionary aspects of protein structure and folding. Curr. Op. Str. Biol. 13, 110—114.

Walker, S.I., Davies, P.C.W., 2017. The algorithm of life. J. R. Soc. Interface 10, 20120869.

Wang, Q., Pang, Y.P., 2007. Preference of small molecules for local minimum conformations when binding to proteins. PLoS One 2 (9), e820.

Wei, B.Q., Weaver, L.H., Ferrari, A.M., Matthews, B.W., Shoichet, B.K., 2004. Testing a flexible-receptor docking algorithm in a model binding site. J. Mol. Biol. 337 (5), 1161—1182.

Wolynes, P.G., 2015. Evolution, energy landscapes and the paradoxes of protein folding. Biochimie 119, 218—230.

Xia, Y., Levitt, M., 2004. Simulated protein evolution is sequence and structure space. Curr. Opin. Struct. Biol. 14, 202—207.

Zsoldos, Z., Reid, D., Simon, A., Sadjad, S.B., Johnson, A.P., 2007. eHiTS: a new fast, exhaustive flexible ligand docking system. J. Mol. Graph. Model. 26 (1), 198—212.

Zwanzig, R., Szabo, A., Bagchi, B., 1992. Levinthal's paradox. PNAS 89, 20—22.