Physica A 388 (2009) 851–862

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Protein folding as an evolutionary process

Vivek Sharma [a], Ville R.I. Kaila [a], Arto Annila [a,b,c,*]

[a] *Institute of Biotechnology, University of Helsinki, Finland*
[b] *Department of Biosciences, University of Helsinki, Finland*
[c] *Department of Physics, University of Helsinki, Finland*

ARTICLE INFO

Article history:
Received 1 October 2008
Available online 7 December 2008

Keywords:
Dissipation
Evolution
Energy landscape
Entropy
Kinetics
Natural process
Natural selection
Self-assembly

ABSTRACT

Protein folding is often depicted as a motion along descending paths on a free energy landscape that results in a concurrent decrease in the conformational entropy of the polypeptide chain. However, to provide a description that is consistent with other natural processes, protein folding is formulated from the principle of increasing entropy. It then becomes evident that protein folding is an evolutionary process among many others. During the course of folding protein structural hierarchy builds up in succession by diminishing energy density gradients in the quest for a stationary state determined by surrounding density-in-energy. Evolution toward more probable states, eventually attaining the stationary state, naturally selects steeply ascending paths on the entropy landscape that correspond to steeply descending paths on the free energy landscape. The dissipative motion of the non-Euclidian manifold is non-deterministic by its nature which clarifies why it is so difficult to predict protein folding.

1. Introduction

Protein folding has attracted intellectual interest ever since 1968 when Cyrus Levinthal pointed out the conceptual difficulty [1]: Why do proteins fold much faster than it would appear to take them to find native states among all conceivable conformations? Today Levinthal's paradox is no longer regarded that much of a mystery [2]. In general, protein folding is pictured to direct down along paths on a free energy funnel landscape, rather than sampling conformational space randomly [3–7]. Conformational entropy of a polypeptide is viewed to decrease along with increasing entropy of the surrounding solvent [8,9] in agreement with the understanding that the *total* entropy must be increasing.

In this study protein folding is described as a *natural process* [10]. Natural processes are evolutionary courses that direct toward more probable states by dispersal of energy [11,12]. They follow the 2nd law of thermodynamics that was recently given as an equation of motion [13] and associated subsequently with the principle of least action [14]. The general formalism establishes correspondence between increasing entropy and decreasing free energy. It has recently given insight to many puzzling natural phenomena and resulting distributions [15–20]. The results are in agreement with earlier findings based on the maximum entropy principle [21–30]]. Thus, the description of protein folding by the ubiquitous imperative provided by this study is not new as such but brings protein folding within the general theory of evolution by natural selection [31].

Protein folding, from random conformational disorder to hierarchical structural order, when formulated properly, follows also the general principle of increasing entropy without an *ad hoc* exemption that entropy of the polypeptide chain would

* Corresponding address: Institute of Biotechnology, University of Helsinki, POB 65, FIN-00014 Espoo, Finland. Tel.: +358 9 191 50629; fax: +358 9 191 50639.
E-mail address: arto.annila@helsinki.fi (A. Annila).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.12.004

be decreasing at the expense of increasing entropy in the surroundings. The new view is in contrast to the common but obscure interpretation that an increase in entropy would invariably coincide with increasing disorder. There is no compelling reason to associate high entropy only with high disorder because entropy $S = k_B \ln P$ is a mere logarithmic measure for the probability P of a state [32,33]. In the light of the 2nd law orderly structures are functional mechanisms of energy dispersal that have emerged during evolution to decrease free energy.

Considering protein folding by the 2nd law of thermodynamics is not only of intellectual interest, but sheds light on the practical folding problem and protein structure prediction. However, it is not claimed that structures could be predicted from amino acid sequences using thermodynamics. On the contrary, it will be clarified that the protein folding is a very hard problem because the evolution of a non-Euclidian energy landscape is non-deterministic and non-integrable [13,14].

2. Protein as a thermodynamic system

It is insightful to picture protein folding as a sequence of chemical reactions using statistical physics. Although covalent bonds, apart from disulphide linkages, do not form during folding, other interactions like hydrogen bonds, ion pairs and also weaker interactions are being established within the polypeptide as well as with and within the surrounding solvent contributing to entropy as hydrophobic effect [34,35]. According to the basic chemical thermodynamics maxim, chemical reactions proceed toward the maximum entropy state, which is equivalent to the minimum free energy state where the chemical potentials of substrates ($\sum \mu_k$) and the product (μ_j) are equal. Protein folding can be formulated in the same way. The simple description will clarify *why* proteins fold, but not *how* specific mechanistic steps lead to the native fold [36].

When a natural process evolves via chemical reactions, it is obvious that the reactants must actually interact with each other for the system to change its state. Energy flows in interactions. Thus, the principle of increasing entropy is a meaningful imperative for a system of interacting entities. This is in contrast with the common misconception that entropy would be a valid concept only for an isolated system accommodating non-interacting or weakly interacting constituents in equilibrium. For the system to change its state, at least a quantum must be expelled to or acquired from the surroundings [13,14]. Of course, one may imagine of isolating an evolving system in a closed space that would accommodate also photons that result from the evolutionary process. However, there is no such a thing as an empty space [37] without any energy, e.g. photons forming the space.

To apply the universal principle of increasing entropy it is immaterial how one wishes to label some energy densities as constituting the system and others as forming the surroundings. Irrespective of the choice the energy differences are diminishing by energy flows that direct down along gradients. Thus, regardless of the viewpoint, entropy of the system, just as entropy of its surroundings, is increasing when free energy is decreasing. The flows connect the system to its surroundings. Therefore the common postulate for the emergence of orderly structures, that the entropy of a system would decrease at the expense of increasing entropy of surroundings, does in fact violate conservation of energy.

For the statistical description of folding, the concept of identity, i.e., ability to distinguish is central [13,14]. Identities are distinguished in interactions. In other words, a hypothetical external observer cannot distinguish polypeptide conformations from each other without interacting with them. First a measurement that couples the observer to the system via a dissipative process [38,39] is able to distinguish one conformation from another. The identity of a functional moiety, e.g., a carboxylic group of an aspartate or a hydroxyl group of a serine, is distinguished in its interactions with other residues and other constituents of the system such as solvent molecules. The change in energy is the measure of distinguishing. In case there is no difference in interaction energy among two moieties, the two are statistically indistinguishable. Of course one may imagine two peptide configurations with identical energy to differ from each other, however, in practice it will turn out to be impossible to separate the inter-converting conformations from each other.

To describe folding as an evolutionary course in the simplest way using the previously described formalism [13,14], the constituents of the initial random coil in numbers N_1 are regarded essentially as indistinguishable substrates with a chemical potential [12] $\mu_1 = RT \ln[N_1 \exp(G_1/RT)]$. The Gibbs free energy G_1 is compared to the average energy RT per mole ($R = k_B N_A$) that includes all other interacting ingredients of the system, e.g., solvent and external fields. Under native conditions the unfolded protein is high in energy compared with its surroundings. Obviously, we realize that the random coil state does not quite make a degenerate pool of conformations but in fact spans a distribution of interaction energies. As long as the band of closely spaced sublevels is narrow in energy there is no significant bias for any particular conformation. Then the diverse random coil conformations distinguish poorly from each other in mutual interactions because their energies are nearly degenerate. This justifies the crude but illustrative approximation of nearly indistinguishable constituents. In this sense, the random coil polypeptide resembles a homopolymer.

Once some mutual interactions, stronger than RT, happen to form, the moieties will begin to distinguish from each other. A folding pathway is opening up in analogy to a reaction pathway opening up (Fig. 1). By pair-forming interactions the substrate pool of N_1 is used in syntheses of products N_2, which can be considered as nascent motifs. Their formation is a way, i.e., a mechanism to disperse energy from the initial random coil state to the surroundings that are lower in energy. Customarily the key residues of nascent motifs that initiate folding are revealed by the Φ-value analysis [40,41]. The chemical potential of products N_2 is $\mu_2 = RT \ln[N_2 \exp(G_2/RT)]$. The pool of N_2 may, in turn, act as the substrates to drive further assembly by energy dispersal to subsequent structural motifs. Differences in the chemical potentials drive the syntheses of larger and larger assemblies (Fig. 1). The number of interactions and their strengths are increasing during the evolutionary