Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 21 (2017) 56–71

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Review

There and back again: Two views on the protein folding puzzle

Alexei V. Finkelstein [a,*], Azat J. Badretdin [b], Oxana V. Galzitskaya [a], Dmitry N. Ivankov [a,c,d], Natalya S. Bogatyreva [a,c,d], Sergiy O. Garbuzynskiy [a]

[a] *Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russian Federation*
[b] *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA*
[c] *Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, 08003 Barcelona, Spain*
[d] *Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain*

Received 5 September 2016; received in revised form 5 January 2017; accepted 19 January 2017
Available online 27 January 2017
Communicated by E. Shakhnovich

Abstract

The ability of protein chains to spontaneously form their spatial structures is a long-standing puzzle in molecular biology. Experimentally measured folding times of single-domain globular proteins range from microseconds to hours: the difference (10–11 orders of magnitude) is the same as that between the life span of a mosquito and the age of the universe. This review describes physical theories of rates of overcoming the free-energy barrier separating the natively folded (N) and unfolded (U) states of protein chains in both directions: "U-to-N" and "N-to-U". In the theory of protein folding rates a special role is played by the point of thermodynamic (and kinetic) equilibrium between the native and unfolded state of the chain; here, the theory obtains the simplest form. Paradoxically, a theoretical estimate of the folding time is easier to get from consideration of protein unfolding (the "N-to-U" transition) rather than folding, because it is easier to outline a good unfolding pathway of any structure than a good folding pathway that leads to the stable fold, which is yet unknown to the folding protein chain. And since the rates of direct and reverse reactions are equal at the equilibrium point (as follows from the physical "detailed balance" principle), the estimated folding time can be derived from the estimated unfolding time. Theoretical analysis of the "N-to-U" transition outlines the range of protein folding rates in a good agreement with experiment. Theoretical analysis of folding (the "U-to-N" transition), performed at the level of formation and assembly of protein secondary structures, outlines the upper limit of protein folding times (i.e., of the time of search for the most stable fold). Both theories come to essentially the same results; this is not a surprise, because they describe overcoming one and the same free-energy barrier, although the way to the top of this barrier from the side of the unfolded state is very different from the way from the side of the native state; and both theories agree with experiment. In addition, they predict the maximal size of protein domains that fold under solely thermodynamic (rather than kinetic) control and explain the observed maximal size of the "foldable" protein domains.

Keywords: Protein folding; Levinthal's paradox; Folding funnel; Phase separation; Free-energy landscape; Protein secondary structure assembly

* Corresponding author.
E-mail address: afinkel@vega.protres.ru (A.V. Finkelstein).

http://dx.doi.org/10.1016/j.plrev.2017.01.025
1571-0645/

1. Introduction

The ability of proteins to fold spontaneously puzzled protein science for a long time (see, e.g., [16,20,50,55,81,98,102]). Our previous review published in PLREV [26] encompassed fundamental experimental facts forming a physical basis of this process and protein physics in general. An updated and extended overview of these facts one can find in a book [30].

It is well known that a protein chain (actually, the chain of a globular protein) can spontaneously fold into its unique native 3D structure [2,3]. In doing so, the protein chain has to find its native (and seemingly the most stable) fold among zillions of others within only minutes or seconds given for its folding.

Indeed, the number of alternatives is vast [62,63]: it is at least 2^{100} but may be 3^{100} or even 10^{100} (or 100^{100}) for a 100-residue chain, because at least 2 ("right" and "wrong"), but more likely 3 (α, β, "coil") or 10 [76] (or even 100 [63]) conformations are possible for each residue. Since the chain cannot pass from one conformation to another faster than within a picosecond (the time of a thermal vibration), the exhaustive search would take at least $\sim 2^{100}$ picoseconds (or 3^{100} or even 10^{100} or 100^{100}), that is, $\sim 10^{10}$ (or 10^{25} or even 10^{80} or 10^{180}) years. And it looks like the sampling has to be really exhaustive, because the protein can "feel" that it has come to the stable structure only when it hits it precisely, while even a 1 Å deviation can strongly increase the chain energy in the closely packed globule.

Then, how does the protein choose its native structure among zillions of possible others, asked Levinthal [62,63] (who first noticed this paradox), and answered: It seems that the protein folding follows some specific pathway, and the native fold is simply the end of this pathway, no matter if it is the most stable chain fold or not. In other words, Levinthal suggested that the native protein structure is determined by kinetics rather than stability and corresponds to the easily accessible local free energy minimum rather than the global one.

However, computer experiments with lattice models of protein chains strongly suggest that the chains fold to their stable structure, i.e., that the "native protein structure" is the lowest-energy one, and protein folding is under thermodynamic rather than kinetic control [1,83].

Nevertheless, most of hypotheses on protein folding are based on the "kinetic control assumption".

Ahead of Levinthal, Phillips [73] proposed that the protein folding nucleus is formed near the N-end of the nascent protein chain, and the remaining chain wraps around it. However, successful *in vitro* folding of many single-domain proteins and protein domains does not begin from the N-end [48,49,60].

Wetlaufer [100] hypothesized formation of the folding nucleus by adjacent residues of the protein chain. However, *in vitro* experiments show that this is not always so [38,99].

Ptitsyn [77] proposed a model of hierarchical folding, i.e., a stepwise involvement of different interactions and formation of different folding intermediate states.

More recently, various "folding funnel" models [4,15,61,98,103] have become popular for illustrating and describing fast folding processes.

The difficulty of the "kinetics vs. stability" problem is that it hardly can be solved by direct experiment. Indeed, suppose that a protein has some structure that is more stable than the native one. How can we find it if the protein does not do so itself? Shall we wait for $\sim 10^{10}$ (or even $\sim 10^{180}$) years?

On the other hand, the question as to whether the protein structure is controlled by kinetics or stability arises again and again when one has to solve practical problems of protein physics and engineering. For example, in predicting a protein's structure from its sequence, what should we look for? The most stable or the most rapidly folding structure? In designing a protein *de novo*, should we maximize stability of the desired fold, or create a rapid pathway to this fold?

However, is there a real contradiction between "the most stable" and the "rapidly folding" structure? Maybe, the stable structure *automatically* forms a focus for the "rapid" folding pathways, and therefore it is *automatically* capable of fast folding?

Before considering these questions, i.e., before considering the *kinetic* aspects of protein folding, let us recall some basic experimental facts concerning protein *thermodynamics* (as usual, we will consider single-domain proteins only, i.e., chains of $\sim$100 residues). These facts will help us to understand what chains and what folding conditions we have to consider. The facts are as follows:

1. The denatured state of proteins, at least that of small proteins treated with a strong denaturant, is often the unfolded random coil [93].