Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Biologies 328 (2005) 701–712
COMPTES RENDUS BIOLOGIES
http://france.elsevier.com/direct/CRASS3/

Review / Revue

Downhill protein folding: evolution meets physics

Martin Gruebele

Departments of Chemistry and Physics, and Center for Biophysics and Computational Biology, University of Illinois, Urbana, IL 61801, USA

Received 4 January 2005; accepted after revision 23 February 2005

Available online 1 April 2005

Presented by Stuart Edelstein

Abstract

Proteins can be redesigned to fold downhill on a free energy surface characterized by only a few coordinates, confirming a principal prediction of the 'energy-landscape' model. Nonetheless, natural proteins have small but significant barriers. Spectroscopy and kinetics reveal potential biological causes for activation barriers during protein folding: evolution against protein aggregation and for protein function. ***To cite this article: M. Gruebele, C. R. Biologies 328 (2005).***

Résumé

Repliement descendant des protéines : l'évolution rencontre la physique. Les protéines peuvent être génétiquement modifiées pour se replier sans barrières signifiantes sur une surface d'énergie libre avec un nombre limité de coordonnées, confirmant une prévision principale du modèle de « paysage d'énergie ». Pourtant, les protéines naturelles ont des barrières petites, mais significatives. Les études cinétiques et spectroscopiques indiquent des causes biologiques potentielles pour les barrières d'activation pendant le repliement des protéines : l'évolution contre l'agrégation des protéines et en faveur de leur fonction. ***Pour citer cet article : M. Gruebele, C. R. Biologies 328 (2005).***

Keywords: Protein function; Temperature jump; Hydrophobicity; Activation barrier

Mots-clés : Fonction de la protéine ; Saut de température ; Hydrophobicité ; Barrière d'activation

1. Introduction and key concepts

Thermodynamically favored reactions of small organic molecules, such as combustion, are generally quite slow at room temperature. They must proceed over large activation barriers during bond-breaking and -making. Protein folding is generally much less favored thermodynamically (protein function often requires proteins to be flexible and at the brink of stability), yet folding is fast at room temperature. In the test tube, denatured states of natural proteins last only for milliseconds to hours under conditions favorable

E-mail address: gruebele@scs.uiuc.edu (M. Gruebele).

1631-0691/$ – see front matter
doi:10.1016/j.crvi.2005.02.007

for folding, in contrast to the long shelf life of organic compounds.

So-called 'water-soluble' globular proteins really fold in a crowded cellular environment in vivo; the largest ones are aided out of misfolded states by chaperones. Yet these proteins unfold and refold spontaneously many times during their lifecycle, and simple mass-action considerations show that cells do not contain enough chaperones to take care of all folding [1]: hence Christian Anfinsen's seminal discovery that the amino acid sequence generally suffices to guide folding of small proteins or protein domains [2], after ribosomal synthesis is complete and without helper-molecules.

The high speed of protein folding, compared to most barrier-controlled chemical reactions, is due to the near-cancellation of enthalpic and entropic contributions to the free energy during the folding process. Proteins can make energy-lowering contacts and become compact in small steps, so no large mismatch appears en route to the folded product. Small barriers in the free energy of folding are distributed along several reaction coordinates, rather than being lumped into one local high-energy barrier. Energy-landscape theory, a statistical-mechanical treatment of protein folding, predicts that this cancellation could be nearly perfect [3]. Such proteins would fold downhill in free energy, on timescales as short as about 0.5 μs for a bundle of three helices.

Natural proteins are not quite that fast, but could proteins be engineered to verify that downhill folding is possible? Fig. 1 shows that the smallest and fastest known folders indeed accomplish the job in about a microsecond. There is kinetic experimental evidence that the folding rate of these fastest folders is limited only by a slight roughness of the free energy surface, with a root-mean-square value $\delta G \approx 1RT \approx 2.5\ \mathrm{kJ\,mol^{-1}}$ [4].

Since downhill-folding proteins can be engineered, a transition state barrier is not a physicochemical requirement for the folding process. What then about the majority of proteins in Fig. 1, whose folding rates lie below the speed limit? Such proteins are said to be 'energetically frustrated' [3]. In addition to the speed limit set by the purely topological requirements of matching up multiple elements of secondary structure in key tertiary contacts, their speed is hampered by non-native contacts and changes in protein–solvent interactions, such as squeezing water molecules out of the hydrophobic core. Such undesirable interactions (from the vantage point of efficient folding) create roughness on the energy landscape.

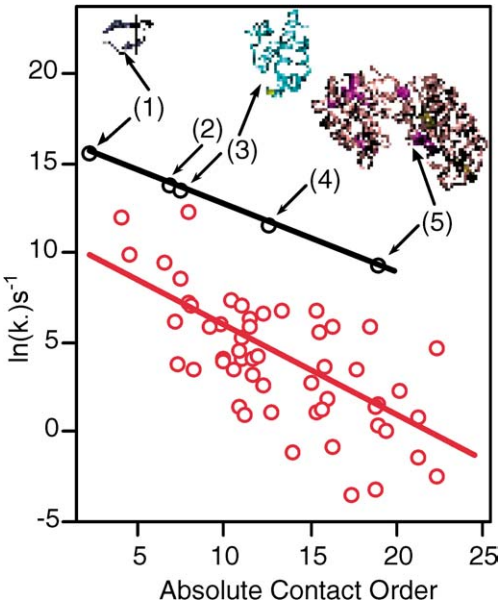

Fig. 1. Correlation of the folding rate with fold complexity (quantified by absolute contact order [50] and illustrated by three folds of increasing complexity). The red line shows the average logarithm of the rate for natural proteins and constructs not specifically engineered for speed from [50]. The black line estimates the folding speed limited only by fold complexity; it decreases exponentially with increasing fold complexity [7]. An alternative model based on homopolymer theory posits a 'softer' linear decrease of the speed limit with sequence length [36]. The molecular rate k_m leading to the native state has been observed directly for λ_{6-85} (3) [14] and for an engineered WW domain [24]. Other speed limit candidates include a single helix (1) [51], the three helix bundle α-3D (2) [26], and the 20 residue trp-cage, observed at 4 μs, with a speed limit probably near 0.5 μs based on our plot [52]. Speed limits estimated from fast-forming intermediates include apomyoglobin (4) [46] and phosphoglycerate kinase (5) [28].

If barriers are not inherently required by the physics of folding, perhaps their roots are to be found in constraints imposed by evolution [5]. Four such constraints, resulting from the interplay of physics with evolution of the amino acid code, of the protein synthesis machinery, for protein function, and against protein aggregation, are considered here.

1. The genetic code evolved early from RNA-peptide interactions, but it is now nearly 'frozen'. Natural proteins are made of 20 natural amino acids,