Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 21 (2017) 72–74

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

Protein folding: Over half a century lasting quest Comment on "There and back again: Two views on the protein folding puzzle" by Alexei V. Finkelstein et al.

Andrey Krokhotin, Nikolay V. Dokholyan *

Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Received 1 June 2017; accepted 1 June 2017
Available online 14 June 2017
Communicated by E. Shakhnovich

Most proteins fold into unique three-dimensional (3D) structures that determine their biological functions, such as catalytic activity or macromolecular binding. Misfolded proteins can pose a threat through aberrant interactions with other proteins leading to a number of diseases including Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis [1,2]. What does determine 3D structure of proteins? The first clue to this question came more than fifty years ago when Anfinsen demonstrated that unfolded proteins can spontaneously fold to their native 3D structures [3,4]. Anfinsen's experiments lead to the conclusion that proteins fold to unique native structure corresponding to the stable and kinetically accessible free energy minimum, and protein native structure is solely determined by its amino acid sequence. The question of how exactly proteins find their free energy minimum proved to be a difficult problem. One of the puzzles, initially pointed out by Levinthal, was an inconsistency between observed protein folding times and theoretical estimates. A self-avoiding polymer model of a globular protein of 100-residues length on a cubic lattice can sample at least 10^{47} states. Based on the assumption that conformational sampling occurs at the highest vibrational mode of proteins (∼picoseconds), predicted folding time by searching among all the possible conformations leads to $\sim 10^{27}$ years (much larger than the age of the universe) [5]. In contrast, observed protein folding time range from microseconds to minutes. Due to tremendous theoretical progress in protein folding field that has been achieved in past decades, the source of this inconsistency is currently understood that is thoroughly described in the review by Finkelstein et al. [6].

Finkelstein and colleagues provide a summary of existing protein folding theories with a special emphasis on the theoretical evaluation of the rates to overcome the free-energy barrier separating the natively folded (N) and unfolded states (U). The authors first make estimates of the free-energy barrier height from the perspective of the folded state by looking into protein unfolding (N-to-U transition). They argue that due to the principle of detailed balance (i.e. the rates of direct and reverse reactions should be equal), the protein folding time can be derived from the time the protein takes to unfold. Then the authors look into the free-energy barrier height from the perspective of the unfolded state (U-to-N transition) and provide rough estimates of the size of the conformational space that the protein samples *en route* to its native structure. This size is significantly lower than the initial estimate suggested by Levintal, hence

DOI of original article: http://dx.doi.org/10.1016/j.plrev.2017.01.025.
* Corresponding author.
E-mail address: dokh@email.unc.edu (N.V. Dokholyan).

http://dx.doi.org/10.1016/j.plrev.2017.06.002

suggesting that there are specific mechanisms that lead to protein folding to its native state bypassing most of the conformational space [7]. Comparing the estimates of the folding times derived from N-to-U and U-to-N transitions, the authors conclude that they are essentially the same and match experimental data. This convergence of the results from two different approaches demonstrates the validity of the described theoretical methods.

The review by Finkelstein and colleagues, in addition to being informative, is written in simple, easy to follow language. Protein folding is a complicated theoretical problem and the authors focus on physical intuition of the readers and minimize the use of mathematics, thus making the review available to a broader audience. One concept that may be especially important to such a broader audience is the volcano-shaped free energy folding landscape. Generally known concept of the free energy folding funnel [8,9] assumes that proteins fold from a denatured state (maximum of the free energy) to a native state (minimum of the free energy) following a smooth energy funnel. This funnel concept, however, cannot explain all-or-none transitions between folded and unfolded states, which assume that a high free-energy barrier separates the native and unfolded states. Whereas the volcano-shaped landscape suggests that during folding, a protein first climbs to the rim of volcano (maximum of the free energy landscape) and only then falls into the volcanos throat (minimum of the free energy landscape).

Besides important protein field overview, it is important to mention a wealth of information obtained from molecular dynamics (MD) simulations of protein folding [10–12]. The last decades were marked by an explosive growth in available computing power and continuous refinements of existing force fields [13,14]. Of the most notable developments worth mentioning is growing use of GPUs (that enables efficient parallelization of MD simulations), harnessing power of distributed computing (like Folding@home project [15]), and custom-built supercomputers for MD simulations (Anton [16]). With these advances, it is currently possible to model the folding of ∼100 residue length protein on the milliseconds timescale using accurate all-atom force fields [17–19]. The accuracy of predicted structures can reach ∼1 Å RMSD (Root Mean Square Deviation) as compared to structures from the Protein Data Bank. Even longer folding time scales can be assessed using MD simulations with coarse-grained force fields [20,21]. The high accuracy of predicted structures and the ability to assess time scales relevant to protein folding, makes MD an invaluable tool for verification of different folding theories. While experiments could be challenging to design and they often do not provide the time resolution necessary to delineate different folding pathways, MD trajectories are readily available and they can be used to analyze protein folding at any relevant time scale (from femtoseconds to milliseconds).

Protein folding field has significantly transformed from focusing on questions pertaining to the basic mechanisms of protein folding to application of these mechanisms for protein design and control of protein function [22]. Finkelstein and colleagues provide a comprehensive summary of the studies done to understand protein folding mechanisms.

References

[1] Valastyan JS, Lindquist S. Mechanisms of protein-folding diseases at a glance. Dis Models Mech 2014:9–14. http://dx.doi.org/10.1242/dmm.013474.

[2] Redler RL, Dokholyan NV. Chapter 7 – the complex molecular biology of Amyotrophic Lateral Sclerosis (ALS). In: Teplow DB, editor. Mol Biol Neurodegener Dis, vol. 107. Academic Press; 2012. p. 215–62.

[3] Anfinsen CB, Haber E, Sela M, White FH. The kinetics of formation of native ribonuclease. Proc. Natl. Acad. Sci. 1961;47:1309–14.

[4] Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181(4096):223–30.

[5] Dokholyan NV. Protein designability and engineering. Struct Bioinforma 2009:963–85.

[6] Finkelstein AV, Badretdin AJ, Galzitskaya OV, Ivankov DN, Bogatyreva NS, Garbuzynskiy SO. There and back again: two views on the protein folding puzzle. Phys Life Rev 2017;21:56–71. http://dx.doi.org/10.1016/j.plrev.2017.01.025 [in this issue].

[7] Šali A, Shakhnovich E, Karplus M. How does a protein fold. Nature 1994;369:248–51.

[8] Leopold PE, Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. Proc. Natl. Acad. Sci. 1992;89:8721–5.

[9] Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science 1995;267(5204):1619–20. http://dx.doi.org/10.1126/science.7886447.

[10] Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253:694–8.

[11] Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–52. http://dx.doi.org/10.1038/nsb0902-646.

[12] Scheraga HA, Khalili M, Liwo A. Protein-folding dynamics: overview of molecular simulation techniques. Annu Rev Phys Chem 2007;58:57–83. http://dx.doi.org/10.1146/annurev.physchem.58.032806.104614.

[13] Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci 2013;3:198–210.

[14] Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 2016.

[15] Beberg AL, Ensign DL, Jayachandran G, Khaliq S. Folding@ home: lessons from eight years of volunteer distributed computing. In: Parallel Distrib. Process. 2009. IPDPS 2009. IEEE Int. Symp.; 2009. p. 1–8.
[16] Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, et al. Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 2008;51:91–7.
[17] Voelz VA, Bowman GR, Beauchamp K, Pande VS. Molecular simulation of ab initio protein folding for a millisecond folder NTL9 (1-39). J. Am. Chem. Soc. 2010;9:1526–8.
[18] Lindorff-larsen K, Shaw DE. How fast-folding proteins fold. Science 2011;334(6055):517–20.
[19] Shirvanyants D, Ding F, Tsao D, Ramachandran S, Discrete NV. Molecular dynamics: an efficient and versatile simulation method for fine protein characterization. J Phys Chem B 2012:8375–82.
[20] Proctor EA, Dokholyan NV. ScienceDirect applications of discrete molecular dynamics in biology and medicine. Curr Opin Struct Biol 2016;37:9–13. http://dx.doi.org/10.1016/j.sbi.2015.11.001.
[21] Ding F, Dokholyan NV. Simple but predictive protein models. Trends Biotechnol 2005;23:450–5.
[22] Chen Y, Ding F, Nie H, Serohijos AW, Sharma S, Wilcox KC, et al. Protein folding: then and now. Arch Biochem Biophys 2008;469:4–19. http://dx.doi.org/10.1016/j.abb.2007.05.014.