COMMENTARY

A simple theoretical model goes a long way in explaining complex behavior in protein folding

Victor Muñoz[a,b,1]

[a]National Biotechnology Center, Consejo Superior de Investigaciones Cientificas, 28049 Madrid, Spain; and [b]School of Engineering, University of California, Merced, CA 95343

Understanding how natural proteins fold spontaneously onto their specific, biologically functional 3D structures is both a fascinating fundamental problem in modern biochemistry and a necessary step toward developing technologies for protein engineering and designing protein-based nanodevices. One of the limitations that scientists working in this area have encountered in the past, however, has been the difficulty in connecting analytical theory to experimental results. For a long time experimentalists could not use theory to interpret their results. Theoretical predictions, moreover, were not amenable to experimental testing. Such limitations have been progressively eliminated by the combination of key theoretical concepts, improved simulations, and new experiments and their detailed quantitative analysis with simple statistical mechanical models. The work of Inanami et al. in PNAS (1) provides a remarkable example of how powerful these simple theoretical models can be in explaining the complexities and nuances of protein folding reactions.

The first major step toward connecting theory and simulations to experiments was initiated by the development of ultrafast kinetic techniques, which led to the experimental determination of the relevant timescales of elementary processes in protein folding such as secondary structure formation and hydrophobic collapse, as well as the identification of several small proteins that fold rapidly (in microseconds) (2). It then became possible to obtain experimental estimates of the folding speed limit (3), a parameter that is essential for interpreting experiments in the context of energy landscape theory (equation 10 in ref. 4). Based on these estimates, the thermodynamic analysis of experimental protein-folding rates revealed that the free-energy barriers to protein folding are indeed entropic bottlenecks (5), as postulated by theory (4). Work on fast-folding proteins also led to the experimental identification of downhill folding (6), a bona fide prediction from energy landscape theory that has encountered tremendous resistance by some experimentalists within the protein-folding community. A second important contribution came from technological developments in computer simulations, which by either distributed computing (7) or computers hard wired for all atom molecular dynamics simulations (8) increased the timescales of atomistic simulations to the point of reaching the folding times of fast-folding proteins, thus permitting the comparison of simulations and experiments on equal footing.

The ability to compare the wealth of structural information included in atomistic simulations with the reality checks provided by experiments is undoubtedly a very exciting development. However, computer simulations, no matter how realistic, cannot substitute analytical theory in interpreting protein-folding experiments. Such a void has been filled through the development and utilization of simple statistical mechanical models of protein folding which permit the direct analysis and fitting of experimental data. In particular, the Ising-like binary model developed in first instance by Wako and Saito (9), and later independently by Muñoz and Eaton to explain folding rates and two-state behavior (10), has proven to be a powerful player in that role (Fig. 1). By far the most important ingredient for the model is that only contacts between residues present in the native structure are attractive, as in the perfect funnel of Onuchic and Wolynes (11). In its original formulation, the Wako–Saito–Muñoz–Eaton (WSME) model described the formation of native structure as the interplay between formation of local nuclei and their growth by propagation through the polypeptide chain. A first major success of this model was its ability to interpret the complex kinetics of the helix–coil transition as well as the simple, two-state-like-folding kinetics of a beta-hairpin (the C-terminal hairpin of protein GB1) (12). The latter was a really surprising result because, in contrast to the essentially 1D process of helix formation, formation of a beta-hairpin already includes all of the elements of a complete folding reaction, such as the competition between local ordering of secondary structure and collapse to form tertiary interactions. The ability to explain beta-hairpin formation was thus a clear hint that the WSME model might be applicable to entire proteins. The model proposed

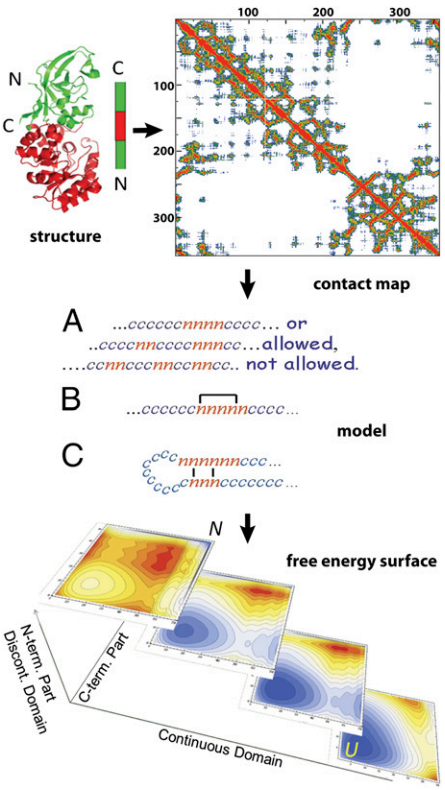

Fig. 1. Ising-like models describe protein-folding free-energy landscapes using the native 3D structure as main input (*Top Left*). The energy function is described by the set of native interactions as shown in the contact map obtained from the 3D structure (*Top Right*) in which all contacts have exactly the same energy. The model defines protein conformations as combinations of residues in native (*n*) and nonnative conformation (coil, *c*) (*Middle*). The entropy loss in the transition from *c* to *n* is assumed to be the same for every residue. The total number of possible conformations can then be simplified by assuming that no more than two contiguous sequences of residues are allowed in each molecule (*A*). In the standard WSME model native contacts occur only if all intervening residues are *n* (*B*). The model can be expanded including the possibility that two native segments interact while connected by a disordered loop (*C*). The projection of the energy and entropy functions onto relevant order parameters leads to the free-energy surface or landscape (*Bottom*).

Author contributions: V.M. wrote the paper.

The author declares no conflict of interest.

See companion article on page 15969.

[1]Email: vmunoz3@ucmerced.edu.

a specific mechanism by which the hairpin folds locally from the turn followed by a zipping up of the strands, and was able to make detailed predictions on the outcomes of further experiments, setting the stage for hundreds of subsequent computational studies using beta-hairpin formation as a benchmark.

The success of the WSME model in predicting folding rates of single-domain proteins from their native 3D structure further confirmed the significance of this simple theoretical approach (10). However, possibly as important was the fact that the analysis provided a simple explanation for the empirical correlations observed between folding rate and protein topology (13) or protein size (14). The WSME model was also pivotal for the analysis of the complex equilibrium properties observed in multiscale studies of one-state downhill folding (6). In subsequent work, an exact analytical solution to the WSME model was obtained by Bruscolini and Pelizzola (15), which simplified calculations and expanded the utility of the model as a unique tool for the quantitative analysis of folding equilibrium and kinetic data on single-domain proteins.

One of the criticisms that the WSME model has encountered over the years is that the mechanism does not allow for the formation of tertiary interactions by closure of disordered loops. It has been often argued that protein chains are not as stiff as the nucleation–elongation mechanism requires, and thus that natural proteins are more likely to fold via a hydrophobic collapse mechanism. To address this criticism, loop-closure was explicitly added to the WSME model by allowing two segments of native structure to interact while separated by a disordered segment (16). Eaton and coworkers used this expanded version of the model to analyze in great depth the equilibrium and folding kinetics of the ultrafast folding villin headpiece subdomain (16). They then compared the folding pathways predicted by the fitted model with the results from long-timescale, all-atom simulations performed by the Shaw group (17). The comparison showed that the folding pathways observed in atomistic simulations involved local nucleation of native structure on no more than two regions, followed by the growth or closure of a single loop. Therefore, the atomistic simulations demonstrated the feasibility of the folding mechanisms invoked by the WSME model.

In retrospect, it is really striking that such a simple model of protein folding could explain so much. However, the results of Inanami et al. (1) take the model to an even higher level of performance. In this case the goal was to analyze the complex folding process of multidomain proteins. To do so, Inanami et al. enhanced the model by cleverly introducing the loop-closure mechanism in the exact analytical solution of the WSME partition function. The role of proline isomerization was also included, which is important to describe the multiphasic kinetics often observed in multidomain protein folding. Armed with an extended WSME model, they tackled the analysis of the folding reaction of dihydrofolate reductase (DFHR), a two-domain protein previously studied experimentally by Matthews and coworkers (18). The experiments reveal a remarkably complex process with up to seven kinetic phases with timescales spanning over 6 orders of magnitude that reflect the formation of early intermediates combined with the slow isomerization of 10 proline residues. DHFR is also interesting from a structural point of view because one of its domains is inserted in the middle of the sequence corresponding to the other domain, thus becoming topologically challenged.

The calculations on DHFR performed with the extended WSME model yielded a rich free-energy landscape with local minima that correspond to the folding of either the central or distal domains (the latter can only form by loop closure). Moreover, the simulation of DHFR-folding kinetics via a Monte Carlo scheme produced multiphasic kinetics with a similar separation of timescales, and more importantly, showed a sequential folding pathway in which the central domain forms first followed by the folding of the distal domain, even though the latter intermediate is thermodynamically more stable. Therefore, the model predicts a kinetically controlled pathway in which the distal intermediate only forms through partial unfolding from the native state. The results obtained with the WSME model offer an interesting rationalization of the experimental information available on DHFR. The model also neatly explains the experiments performed on a circular permutation of DHFR in which the elimination of the topological complexity seems to rebalance the flux between two alternative folding pathways, making them equally populated.

The significance of the results of Inanami et al. (1) goes well beyond the success in reproducing DHFR folding. The existing tools for the analysis of multidomain-folding process pale in comparison with the sophisticated experimental and computational procedures now available for investigating fast-folding proteins. Therefore, the extension of the WSME model to the analysis of multidomain folding and its ability to reproduce experimental results and recapitulate the conclusions of coarse-grained computer simulations of DHFR folding (19) are indeed excellent news. It gives us hope for the near future in which simple statistical mechanical models may become extensively used for the analysis of complex protein-folding experiments, thus mimicking the pivotal role they have played in the study of fast-protein folding.

1 Inanami T, Terada TP, Sasai M (2014) Folding pathway of a multidomain protein depends on its topology of domain connectivity. *Proc Natl Acad Sci USA* 111:15969–15974.

2 Eaton WA, Muñoz V, Thompson PA, Chan CK, Hofrichter J (1997) Submillisecond kinetics of protein folding. *Curr Opin Struct Biol* 7(1):10–14.

3 Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit' *Curr Opin Struct Biol* 14(1):76–88.

4 Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21(3):167–195.

5 Akmal A, Muñoz V (2004) The nature of the free energy barriers to two-state folding. *Proteins: Struct, Funct, Bioinf* 57(1):142–152.

6 Garcia-Mira MM, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V (2002) Experimental identification of downhill protein folding. *Science* 298(5601):2191–2195.

7 Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* 420(6911):102–106.

8 Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517–520.

9 Wako H, Saito N (1978) Statistical mechanical theory of protein conformation 2. folding pathway for protein. *J Phys Soc Jpn* 44(6): 1939–1945.

10 Muñoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96(20):11311–11316.

11 Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14(1):70–75.

12 Muñoz V, Thompson PA, Hofrichter J, Eaton WA (1997) Folding dynamics and mechanism of beta-hairpin formation. *Nature* 390(6656):196–199.

13 Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277(4):985–994.

14 Naganathan AN, Muñoz V (2005) Scaling of folding times with protein size. *J Am Chem Soc* 127(2):480–481.

15 Bruscolini P, Pelizzola A (2002) Exact solution of the Muñoz-Eaton model for protein folding. *Phys Rev Lett* 88(25 Pt 1):258101.

16 Kubelka J, Henry ER, Cellmer T, Hofrichter J, Eaton WA (2008) Chemical, physical, and theoretical kinetics of an ultrafast folding protein. *Proc Natl Acad Sci USA* 105(48): 18655–18662.

17 Henry ER, Best RB, Eaton WA (2013) Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. *Proc Natl Acad Sci USA* 110(44):17880–17885.

18 Smith VF, Matthews CR (2001) Testing the role of chain connectivity on the stability and structure of dihydrofolate reductase from E. coli: Fragment complementation and circular permutation reveal stable, alternatively folded forms. *Protein Sci* 10(1):116–128.

19 Clementi C, Jennings PA, Onuchic JN (2000) How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta. *Proc Natl Acad Sci USA* 97(11):5871–5876.