Behind the folding funnel diagram

Martin Karplus

This Commentary clarifies the meaning of the funnel diagram, which has been widely cited in papers on protein folding. To aid in the analysis of the funnel diagram, this Commentary reviews historical approaches to understanding the mechanism of protein folding. The primary role of free energy in protein folding is discussed, and it is pointed out that the decrease in the configurational entropy as the native state is approached hinders folding, rather than guiding it. Diagrams are introduced that provide a less ambiguous representation of the factors governing the protein folding reaction than the funnel diagram.

Understanding the mechanism by which proteins fold to their native state remains a problem of fundamental interest in biology, in spite of the fact that it has been studied for many years[1]. Moreover, now that misfolding has been shown to be the source of a range of diseases, a knowledge of the factors that determine whether a polypeptide chain will fold to its native state or aggregate has become all the more important.

The folding funnel diagram (see **Fig. 1**), introduced by Wolynes, Onuchic and Thirumalai[2], is intended to provide a pictorial representation of how the Levinthal paradox[3], which had dominated discussion of protein folding for many years, is resolved. Since the original publication, funnel diagrams have become a fixture in papers on protein folding, and they are now being introduced in discussions of other problems, such as ligand binding[4]. Unfortunately, the funnel diagram has created a misconception in many readers.

The concept introduced by Levinthal is that the appropriate point of reference for protein folding is a random search problem. Taken literally, as it has been by many people, this means that all conformations of the polypeptide chain (except the native state) are equally probable, so that the native state can be found only by an unbiased random search. For such a search, the time to find the native state is given by the number of configurations of the polypeptide chain (on the order of 10^{70} for a 100-residue protein) multiplied by the time required to find one configuration (say, 10^{-11} seconds). This leads to an enormously long folding time (say, 10^{59} seconds or about 10^{52} years). Given that small proteins (100 residues or fewer) generally fold in times on the order of milliseconds to seconds (except in cases with special factors that slow the folding, such as proline isomerization), there was indeed a paradox. The ultimate statement of the paradox was given in the language of computational complexity[5]. Many phenomenological models were proposed to show how the conformational space that has to be searched is restricted to reduce the folding time to the experimental range. Examples include the nucleation-growth or nucleation-condensation mechanism[6,7], the diffusion-collision model[8] and the jigsaw-puzzle model[9].

In the late 1980s, the focus of approaches to the protein folding problem shifted from phenomenological models to a consideration of the general characteristics of the energy surface of a polypeptide chain. The new focus is eminently reasonable, as the energy surface is one of the fundamental determinants of any reaction[1], whether a small-molecule reaction or protein folding. The change made explicit a concept that is implicit in the phenomenological models—for any of the models to work, there must be energetic factors that bias the folding process. For example, only if a nucleus is stable, relative to the random coil structures, can it play a role in folding. In an insightful paper, Zwanzig *et al.*[10] used a simplified model to demonstrate that if there is a bias in the potential energy such that it decreases relatively smoothly toward the native state, only a limited number of configurations would be visited in the folding reaction, and the configurational search (Levinthal) problem would be solved.

The required energy bias is embodied in the two-dimensional funnel diagram, the most widely used of which is similar to that shown in **Figure 1**. It is a schematic representation of how the effective potential energy, implicitly averaged over solvent interactions (in the vertical direction), and the configurational entropy (in the horizontal direction) of a protein decrease as the native state is approached; picturesque three-dimensional funnels are also being used[11]. Both the two- and three-dimensional diagrams have a funnel-like shape because the number of accessible configurations, which determine the configurational entropy, decreases as the energy decreases. Such funnel diagrams are very appealing images from which many readers have concluded that folding a protein is like "funneling wine into a bottle." A typical statement of this type appears in a study of the folding of PDZ domains[12]; in the introduction, it is stated that "Free energy landscapes of many proteins appear to resemble the shape of a funnel that *guides* the folding process toward the native state." Although Wolynes *et al.*[2] were aware that the occurrence of

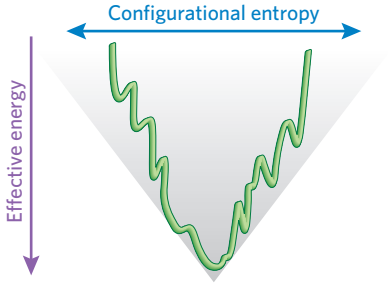

Figure 1 | A schematic folding funnel diagram. The effective energy is plotted vertically and the configurational entropy horizontally. Adapted from reference 2.

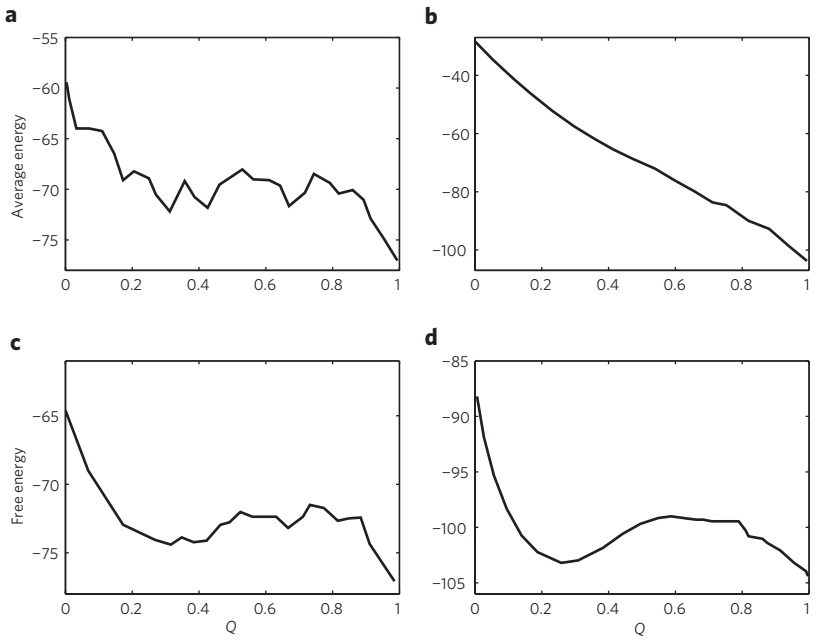

Figure 2 | Folding of a lattice polymer. (**a–d**)The effective energy (**a**,**b**) and free energy (**c**,**d**) calculated as a function of the fraction of native contacts for the 27-mer lattice polymer: **a** and **c** at a low temperature; and **b** and **d** at a high temperature. Adapted with permission from reference 22.

such "guiding" is a misconception, as can be seen from their original paper, they have not emphasized this point, and the misconception concerning the funnel diagram has continued to proliferate.

The essence of the misconception is that the decrease in configuration entropy, which gives the diagram its funnel-like shape, aids the polypeptide chain in finding the native state. In fact, exactly the opposite is true; that is, the decrease in the number of available configurations, as the native state is approached, actually tends to slow folding. The difficulty of finding these configurations is essentially the origin of the Levinthal paradox. To understand the folding process it must be realized that the major determinant of the folding rate is the free-energy surface (often referred to as a "landscape") of the polypeptide chain[1,2,7,13], rather than the energy shown in the funnel diagram. The free energy is the sum of the potential energy, which decreases as the native state is approached and therefore favors folding, and the *unfavorable* contribution of the decrease in the configuration entropy. The delicate balance between the two generally leads to a free-energy barrier that results in the two-state folding behavior observed for most small proteins (see ref. 7 and references therein).

The developing understanding of protein folding has been aided by experiments that provide structural and kinetic information that contribute to dissecting the mechanism, such as those making use of NMR[1] and protein engineering[7]. Also, it has become possible to rapidly trigger folding and unfolding so that measurements can be extended from milliseconds down to microseconds and more recently to nanoseconds[14]. An important result of such studies is the finding that the folding reaction of many small proteins follows an exponential time course. This is interpreted to mean, as in chemical reactions[1], that there are only two significantly populated states, the denatured state and the native state, and that there is a barrier separating the two that is larger than a few kilocalories. However, it has so far not been possible to determine experimentally the detailed folding trajectories even for a simple protein—that is, a complete description of the conformations that are sampled in going from the multiconfiguration denatured state to a well-defined native state. Consequently, a range of theoretical methods continues to be used to supplement the experimental data. Models that go beyond the phenomenological approaches mentioned above by focusing on the structural and thermodynamic parameters include lattice models[1,13], coarse-grained models[15] and atomistic simulations with implicit (see, for example, ref. 16) and explicit (see, for example, ref. 17) representations of the solvent. Also, unfolding simulations in explicit solvent, combined with experimental data, have been used to propose folding pathways for small proteins (see, for example, ref. 7).

The first concrete demonstration of how the Levinthal paradox can be resolved

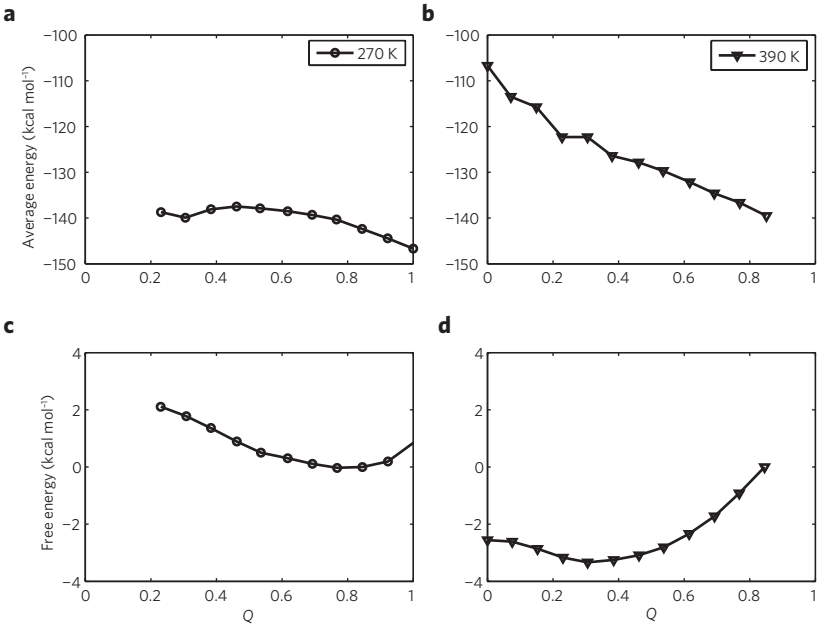

Figure 3 | Folding of a designed α-helical peptide. (**a–d**) The effective energy (**a**,**b**) and free energy (**c**,**d**) calculated as a function of the fraction of native hydrogen bonds for a designed α-helical peptide: **a** and **c** at a low temperature (270 K); **b** and **d** at a high temperature (390 K). Adapted with permission from reference 21; some data points for low and high *Q* values at the low and high temperature, respectively, are not shown because of poor statistics.

was provided by Monte Carlo simulations of a 27-residue polypeptide represented by a chain of beads on a cubic lattice with interactions between neighboring beads favoring the native state[13] (see also ref. 18). The simulations demonstrated that folding occurred even when the number of denatured configurations is large enough (10^{18}) to make it impossible for a random search to find the native state in the simulation time. In accordance with the conceptual framework of the Zwanzig *et al.* model[10], it was shown that only about 10^7 steps were required to fold the sequences, whereas there are on the order of 10^{16} configurations in the denatured state. Analysis of molecular dynamics simulations of the folding of a 20-residue peptide that forms a stable three-stranded double-hairpin β-sheet represented by an all-atom potential energy function showed results in accord with those obtained from the lattice simulations; during the average folding time, the peptide folds to the native state and solves the Levinthal paradox by having to visit only an infinitesimal fraction of the denatured configurations[19].

One-dimensional representations showing the free energy as a function of the distance from the native state provide more information about the folding mechanism and are less subject to misinterpretations than the funnel picture[1,7,13,20]. **Figures 2** and **3** illustrate this point; they are based on the 27-mer lattice Monte Carlo simulation described above[13]. **Figure 2a,c** shows the lattice model results for the effective energy and free energy at a low temperature and **Figure 2b,d** at a high temperature, as a function of the number of native contacts, *Q*, which is a satisfactory progress variable for the system. At the lower temperature (**Fig. 2a**), the effective energy surface resulting from the sampled configurations is 'rugged' and does not have a monotonic decrease as the 27-mer structure approaches the native state. The free energy values (**Fig. 2c**) show that the added configurational entropy introduces a barrier; it is in the region of $Q = 0.8$. For the higher temperature (**Fig. 2b**), in contrast, the effective energy of the sampled configuration decreases smoothly as the native state is approached, whereas the free energy (**Fig. 2d**) has a significant barrier. The latter is responsible for the exponential (two-state) folding kinetics.

Although there are no corresponding all-atom simulations or experimental results showing the free energy and energy for the folding of a protein, all-atom peptide studies in implicit solvent over the appropriate temperature range have been reported for a synthetic α-helix and are shown in **Figure 3** (for details, see Ferrara *et al.*[21]). As with the lattice model, the Figure presents both low- and high-temperature results for the effective energy and free energy as a function of the progress coordinate *Q*, here defined in terms of the fraction of the number of hydrogen bonds in the folded state. The effective energy plots (**Fig. 3a,b**) are similar to those in **Figure 2a,b**, except that the low-temperature result is somewhat less rugged than for the lattice model; the higher-temperature effective energy decreases smoothly to the native state as in **Figure 2b**. As to the free energy (**Fig. 3c,d**), at low temperature there is a stable minimum at $Q = 0.8$, whereas at high temperature the helical state is not stable; a barrier is not clearly visible, though it is likely to occur very early when the first helical turn is formed. In corresponding studies of a somewhat more complex peptide, the three-stranded β-sheet mentioned above[19], the energy and free-energy folding diagrams are more complex. What happens in actual proteins (for example, is **Fig. 1** valid for any protein?) is not known.

The important role of the enthalpy and the entropy in the protein folding reaction is confirmed by the fact that there are striking deviations from the Arrhenius temperature dependence of the folding rate in lattice[13,22] and all-atom models[21] (see **Fig. 4a,b**). The rate as a function of temperature increases at low temperatures, in accord with the Arrhenius formula. At higher temperatures, however, there is a turnover in the rate constant that *decreases* as the temperature is increased[22]; corresponding results have been observed experimentally[7] (see Fig. 18.2 in ref. 7). A possible explanation of this behavior is evident from **Figure 2**. At low temperature (**Fig. 2a**), the folding rate is dominated by the ruggedness of the potential surface that leads to an apparent activation energy. As the temperature increases, the apparent activation energy becomes negative (**Fig. 2b**), whereas the activation free energy remains positive and is dominated by the increasing (unfavorable) entropic contribution because of the larger number of accessible configurations that must be searched to find the native one. Thus, from the physicist's viewpoint, proteins are 'hard matter' at low temperature (enthalpy dominated), whereas they are 'soft matter' (entropy dominated) at high temperature[23]. What corresponds to a 'low' and 'high' temperature is likely to depend on the specific protein.

The funnel concept has stimulated useful studies of protein folding; see, for example, reference 24. I hope that this Commentary will eliminate the misconceptions that the funnel diagram has engendered in future discussions of protein folding. Diagrams like those shown here in **Figures 2–4**, supplemented by figures based on the analysis of molecular dynamics simulations of peptides repeatedly folding and unfolding at equilibrium (near the transition temperature) (see ref. 25 and references therein) provide a more meaningful description of the protein folding reaction. The reference shows that there are hidden complexities not evident from experiment nor present in the folding funnel diagram. ■

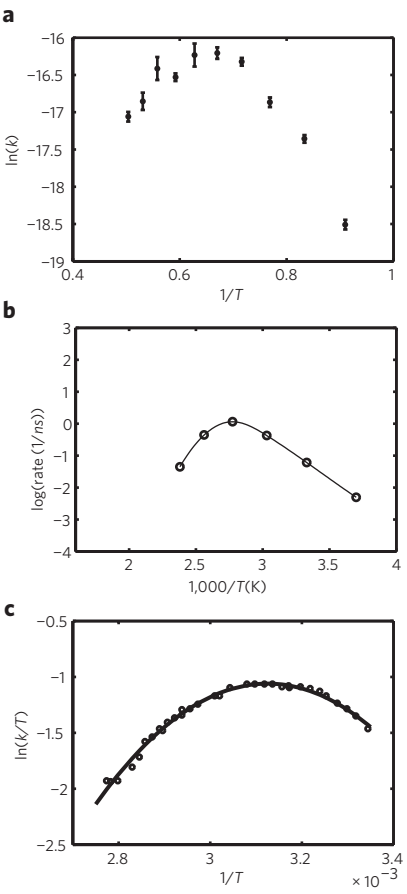

Figure 4 | Arrhenius plots for folding reaction. (**a**) Calculated rate constant as a function of temperature for the lattice polymer in **Figure 2**; adapted with permission from reference 2. (**b**) Calculated rate constant as a function of temperature for the designed α-helical peptide in **Figure 3**. Adapted with permission from reference 21. (**c**) Measured refolding rate constant for CI2. Adapted with permission from references 7 and 22.

Martin Karplus is in the Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts, USA and in the Laboratoire de Chimie Biophysique, Institut de Science et d'Ingénierie Supramoléculaires, Université de Strasbourg, Strasbourg, France.
e-mail: marci@tammy.harvard.edu

References

1. Dobson, C.M., Sali, A. & Karplus, M. *Angew. Chem. Int. Edn. Engl.* **37**, 868–893 (1998).
2. Wolynes, P.G., Onuchic, J.N. & Thirumalai, D. *Science* **267**, 1619–1620 (1995).
3. Levinthal, C. How to fold graciously. in *Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting held at Allerton House, Monticello, Illinois* (Debrunner, P., Tsibris, J.C.M. & Münck, E., eds.) 22 (University of Illinois Press, Urbana, 1969).
4. Tsai, C.J., Kumor, S., Ma, B.Y. & Nussinov, R. *Protein Sci.* **8**, 1181–1190 (1999).
5. Ngo, J.T., Marks, J. & Karplus, M. Computational complexity, protein structure prediction, and the Levinthal paradox. in *The Protein Folding Problem and Tertiary Structure Prediction*. (Merz, K. Jr. & Le Grand, S., eds.), 435–508 (Birkhäuser, Boston, 1994).
6. Wetlaufer, D.B. *Proc. Natl. Acad. Sci. USA* **70**, 697–701 (1973).
7. Fersht, A. *Structure and Mechanism in Protein Science: A guide to Enzyme Catalysis and Protein Folding* (WH Freeman, 1999).
8. Karplus, M. & Weaver, D.L. *Nature* **260**, 404–406 (1976).
9. Harrison, S.C. & Durbin, R. *Proc. Natl. Acad. Sci. USA* **82**, 4028–4030 (1985).
10. Zwanzig, R., Szabo, A. & Bacghi, B. *Proc. Natl. Acad. Sci. USA* **89**, 20–22 (1992).
11. Dill, K.A. & Chan, H.S. *Nat. Struct. Biol.* **4**, 10–19 (1997).
12. Calosci, N. *et al. Proc. Natl. Acad. Sci. USA* **105**, 19241–19246 (2008).
13. Sali, A., Shakhnovich, E. & Karplus, M. *Nature* **369**, 248–251 (1994).
14. Kubelka, J., Henry, E.R., Cellmer, T., Hofrichter, J. & Eaton, W.A. *Proc. Natl. Acad. Sci. USA* **105**, 18655–18662 (2008).
15. Lammert, H., Schug, A. & Onuchic, J. *Bioinformatics* **77**, 881–891 (2009).
16. Zagrovic, B. & Pande, V. *J. Comput. Chem.* **24**, 1432–1436 (2003).
17. Shaw, D.E. *et al. Science* **330**, 341–346 (2010).
18. Baldwin, R.L. *Nature* **369**, 183–184 (1994).
19. Cavalli, A., Haberthuer, U., Paci, E. & Caflish, A. *Protein Sci.* **12**, 1801–1803 (2003).
20. Eaton, W.A., Thompson, P.A., Chan, C.-K., Hagen, S.J. & Hofrichter, J. *Structure* **4**, 1133–1139 (1996).
21. Ferrara, P., Apotolakis, J. & Caflisch, A. *J. Phys. Chem. B* **104**, 5000–5010 (2000).
22. Karplus, M. *J. Phys. Chem. B* **104**, 11–27 (2000).
23. Müller-Plathe, F. Introduction. in *Soft Matter Vol. 1, Polymer Melts and Mixtures* (Gompper, G. & Schick, M., eds.) (Wiley, 2006).
24. Onuchic, J.N., Socci, N.D., Luthey-Schulten, Z. & Wolynes, P.G. *Fold. Des.* **1**, 441–450 (1996).
25. Krivov, S.V., Muff, S., Caflisch, A. & Karplus, M. *J. Phys. Chem. B* **112**, 8701–8714 (2008).

Acknowledgments

I thank A. Caflisch, A. Dinner and B. Qi for providing results and comments on the manuscript and M. Cecchini and V. Ovchinnikov for comments on the manuscript and help with the figures. Owing to space limitation, some original citations have been omitted; many can be found in reference 1 or reference 7. The work done at Harvard was supported, in part, by a grant from the US National Institutes of Health.

Competing financial interests

The author declares no competing financial interests.

Corrected after print 14 July 2011

CORRIGENDUM

Behind the folding funnel diagram

Martin Karplus

Nat. Chem. Biol. **7**, 401–404 (2011); published online 17 June 2011; corrected after print 14 July 2011

In the version of this article initially published, there was an error in the abstract that stated an increase in the configurational entropy hinders folding, but it should read that there is a decrease in the configurational entropy that hinders folding. The error has been corrected in the HTML and PDF versions of the article.