COMMENTARY

Heterogeneous folding and stretched kinetics

Osman Bilsel[a,b,c,1]

It has been over half a century since Anfinsen et al. (1) outlined in a series of papers their groundbreaking discovery that the amino acid sequence of a globular protein determines its native structure. This hypothesis laid the foundation for a thermodynamic view of protein folding, where the native state is energetically favored over all other conformations of the protein. One question left open, however, was the process by which this occurred: How did proteins arrive at their native conformation on a biologically practical timescale? A broad area of kinetic investigations in a 50+-y journey of kinetic studies uncovered a wealth of insights into protein structure, dynamics, and kinetics that has had a tremendous impact on our current understanding of biological machinery and regulation at the molecular level.

Early pioneers in protein science studying how proteins spontaneously transformed from what appeared to be a spaghetti-like self-avoiding random walk chain to a specific native conformation noticed that the process did not necessarily occur in a single kinetic step (2). A reaction with a single kinetic step (i.e., kinetically limited by one dominant energy barrier) exhibits a time course described by a single exponential function. However, the folding kinetics for a number of proteins required two (or more) exponentials to describe the data. This result was both exciting, because of the opportunity for gaining insights into the folding process, and also a source of debate: Did the additional steps correspond to obligatory intermediates along the folding process, as in a chemical reaction, or did they suggest the presence of traps? In kinetic studies, it is often possible for multiple kinetic models to adequately describe the data, and often the experimenter can only rule out inconsistent models. For ensemble experiments, a combination of experimental techniques sensitive to more than a few molecular properties of a system is extremely helpful in this effort and requires careful attention to the details of the amplitudes of the exponentials under a variety of conditions. For example, a fluorescence experiment may be sensitive only to the local environment of tryptophan side chains, whereas circular dichroism (and its counterpart, optical rotary dispersion) may report primarily on secondary structure formation, a global property of the protein chain. A technique simultaneously sensitive to all length scales of the molecule (i.e., local and global conformational changes) would therefore be helpful. Additionally, a time resolution spanning the timescale of fundamental events (3), such as loop formation (which occur in microseconds), helix formation (hundreds of nanoseconds), to native state formation (tens of microseconds to seconds), would be ideal.

In PNAS, Kim et al. (4) apply an experiment that meets all of these criteria to test the kinetic folding model of the "hydrogen molecule of proteins," cytochrome *c*. Folding of oxidized cytochrome *c* has been studied using perhaps every spectroscopic and labeling technique developed for proteins since the early days of protein science (5). Its use stems partly from being commercially available, prepurified, in large quantities [in the early days (6), the horse variant was available from horse hearts provided by slaughterhouses] and is perhaps the most thoroughly studied protein in the folding literature. The reduced form, however, has been much less studied, and the studies to date have not provided a consensus on the time constants or the mechanism describing its folding kinetics (7–9). Although cytochrome *c* is a small (104 amino acids) helical single-domain monomeric protein that has long served as a model system, it is complicated by several factors: It has a centrally located heme moiety, multiple histidine and methionine side chains capable of ligating to the heme, and proline residues that can undergo *cis–trans* isomerization on timescales comparable to or slower than the folding time. The extent to which histidine–heme misligation events affect the folding kinetics is one of the key unanswered questions. Some researchers

[a]Microgradient Fluidics LLC, Worcester, MA 01605; [b]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605; and [c]Department of Chemistry, Stony Brook University, Stony Brook, NY 11794

Author contributions: O.B. wrote the paper.

The author declares no competing interest.

See companion article, "Protein folding from heterogeneous unfolded state revealed by time-resolved X-ray solution scattering," 10.1073/pnas.1913442117.

[1]Email: osman.bilsel@gmail.com.

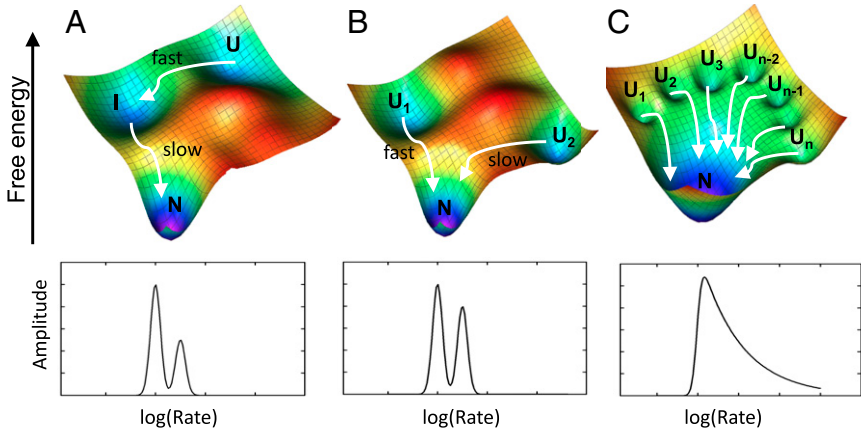

Fig. 1. Cartoon of energy landscapes and their corresponding rate distributions. (*A*) A kinetic mechanism in which an intermediate is populated. Two distinct rates would be observed experimentally (below) with amplitudes dependent on both the relative microscopic rates and the relative signals of the unfolded state (U), intermediate state (I), and native state (N). (*B*) Two distinct rates (*Lower*) would also be observed if two unfolded states (U_1 and U_2) fold independently faster than they can interconvert. (*C*) When microstates in a heterogeneous unfolded ensemble can access the native or compact state faster than they can interconvert, a broad rate distribution is obtained (*Lower*) that gives rise to a "stretched exponential" or "strange kinetics."

have avoided this issue altogether by adjusting the pH (protonating the histidine side chain) or addition of imidazole (outcompeting intramolecular ligation events) to focus on folding in the absence of these complications. In these cases, the folding kinetics are greatly simplified with no evidence of intermediates (7). With the introduction of histidine misligation, however, the sequence of events is potentially much more complicated: There are three histidine and two methionine side chains that can ligate to the heme, and the order in which they ligate, assuming that misligation and the order of misligation matters (7), is tricky to delineate. Kim et al. weigh in to this debate with a tour-de-force experiment that combines fast optically triggered reduction of the heme (to initiate refolding), small-angle X-ray scattering (SAXS), and computational simulations. These methods have been separately employed before (8–10), but the combination provides a unique window into the energy landscape. This is in part because SAXS is sensitive to all pairs of intramolecular distances and contains information on the three-dimensional structure of the protein conformations, albeit low resolution by modern X-ray crystallography and cryogenic electron microscopy standards.

The folding kinetics observed by Ihee and colleagues differs from previous studies in that their data are best described not by one or two exponential decay functions, as might be expected for a reaction with discrete kinetic steps, but by a "stretched" exponential function, $y = a{\cdot}\exp(-t/\tau)^{\beta}$, where τ is a time constant and β is a fractional exponent between 0 and 1. This type of decay was first observed in 1854 by Kohlrausch in his dielectric relaxation experiments in a Leyden jar, in which he was essentially monitoring the discharging of a capacitor (11). He modeled his observations using a modified first-order reaction [$dQ(t)/dt = -\gamma(t)^{*}Q(t)$], where the typically time-invariant proportionality constant, $\gamma(t)$, has an inverse power-law time-dependent transition rate $\gamma(t) = t^{-(1\,-\,\beta)}$ (11). Integration yields a fractional exponent, $0 \leq \beta \leq 1$ [$Q = Q_o{\cdot}\exp(-t/\tau)^{\beta}$], for the modified exponential decay of the charge, and hence the term stretched used in the modern literature (12). As noted in ref. 11, the fractional exponent is also characteristic of Levy statistics generated by random processes that are scale invariant. Highly nonexponential kinetics have been observed in other disordered condensed matter systems. For example, turbulent diffusion where "intermittency" is used to describe traps follows this type of kinetics (11). A physical model corresponds to a particle exhibiting Brownian motion with traps or sinks of fixed radius (11). The results of Kim et al. suggest stretched folding kinetics can originate from heterogeneity in the unfolded state introduced by trapped states, presumably from misligated ligands, other specific or nonspecific intramolecular interactions, or isomerization of proline side chains. The main point is that the trapped states can access the native state directly and independently (7), as illustrated in Fig. 1.

The observations of Kim et al. add to other examples of highly nonexponential behavior in proteins. Pioneering experiments by Austin et al. (13) showed heterogeneous kinetics in low-temperature CO photolysis and rebinding experiments of myoglobin, the other "hydrogen molecule" of proteins. This work played a significant role in popularizing the concept of an energy landscape and illustrating heterogeneous kinetics for proteins. Stretched exponential kinetics of CO rebinding has also illustrated the challenges of experimentally establishing stretched kinetics: Specifically, the absence of early time data and incomplete knowledge of the amplitudes of the initial species can skew the kinetic analysis (e.g., introducing peaks in the rate expression) (14). The fast optically triggered folding used by Kim et al. overcomes both of these limitations. In their studies, the experiment monitors the protein from being nearly fully unfolded to completely folded. This is perhaps a key difference from experiments by the Gruebele group (15), who also observed stretched kinetics (or "strange kinetics") in folding experiments initiated by optically triggered temperature jumps from the cold denatured state. The studies by the Gruebele group, while having excellent temporal resolution, used local probes (e.g., tyrosine and tryptophan fluorescence) and complemented the kinetic studies with mutational analysis to confirm stretched exponential behavior.

The finding of heterogeneity in the unfolded state reinforces an idea that has become increasingly appreciated in protein science. In fact, slow *cis–trans* isomerization of prolyl peptide bonds, for example, which can be slower than the folding rate to the native state, has been known for decades to give rise to heterogeneous kinetics, often displaying as parallel kinetic channels (16). The fast time resolution of optically triggered experiments extends our ability to observe heterogeneity when subensembles

interconvert on a timescale comparable to early events in folding (e.g., collapse, loop formation, and helix formation) (3). The approach is not limited to systems with misligation to a heme, as suggested by the Gruebele group's results, and it can be more broadly applied to uncovering energetic biases in the energy landscape arising from residue-specific interactions in the unfolded state. Surprisingly, slow interconversion of subensembles that do not include proline isomerization or heme misligation has been suggested based on theoretical and computational folding studies. In the kinetic partitioning mechanism (17) and the hub-and-spoke model (18), subensembles in the unfolded state can access compact and native conformations more readily than other unfolded state subensembles, analogous to what Kim et al. observe. The current work is a step toward facilitating experimental tests of this counterintuitive idea. Slow interconversion between unfolded state subensembles may also be relevant in human disease, as suggested by studies by the Lapidus group (19) on the disordered peptide α-synuclein.

In closing, it may be worth noting that despite the tremendous success over the past 50 y in addressing some of the fundamental questions in protein science, the role of amino acid sequence in modulating the conformational dynamics and specific transient interactions within the unfolded ensemble remains a current challenge. These interactions have become increasingly appreciated over the past decade via studies revealing the role of disordered proteins in regulation and cellular function (20). Ultimately, one goal is to obtain a molecular-level description of the unfolded or disordered ensemble. An important component of a strategy toward achieving this is the synergy between experiment and computational simulations (21). Experimental techniques reporting on interatomic distances and their distributions, such as SAXS, are particularly amenable to this. The study by Kim et al., coupling tour-de-force experimental techniques with molecular dynamics simulations, is therefore an important step in bringing us closer to realizing a quantitative atomistic-level understanding of the upper reaches of the energy landscape of proteins.

Acknowledgments

I thank Dave Thirumalai for helpful discussions on the kinetic partitioning mechanism and Jill Zitzewitz for critical review of this commentary.

1 C. B. Anfinsen, E. Haber, M. Sela, F. H. White, Jr, The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1309–1314 (1961).
2 A. Ikai, C. Tanford, Kinetics of unfolding and refolding of proteins. I. Mathematical analysis. *J. Mol. Biol.* **73**, 145–163 (1973).
3 W. A. Eaton *et al.*, Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359 (2000).
4 T. W. Kim *et al.*, Protein folding from heterogeneous unfolded state revealed by time-resolved X-ray solution scattering. *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1913442117 (2020).
5 R. A. Goldbeck, E. Chen, D. S. Kliger, Early events, kinetic intermediates and the mechanism of protein folding in cytochrome *c*. *Int. J. Mol. Sci.* **10**, 1476–1499 (2009).
6 J. Babul, E. Stellwagen, Participation of the protein ligands in the folding of cytochrome *c*. *Biochemistry* **11**, 1195–1200 (1972).
7 A. K. Bhuyan, J. B. Udgaonkar, Folding of horse cytochrome *c* in the reduced state. *J. Mol. Biol.* **312**, 1135–1160 (2001).
8 J. R. Telford, F. A. Tezcan, H. B. Gray, J. R. Winkler, Role of ligand substitution in ferrocytochrome *c* folding. *Biochemistry* **38**, 1944–1949 (1999).
9 D. Rimmerman *et al.*, Probing cytochrome *c* folding transitions upon phototriggered environmental perturbations using time-resolved X-ray scattering. *J. Phys. Chem. B* **122**, 5218–5224 (2018).
10 G. Tria, H. D. T. Mertens, M. Kachala, D. I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).
11 J. C. Phillips, Microscopic theory of the Kohlrausch relaxation constant βK. *J. Non-Cryst. Solids* **172–174**, 98–103 (1994).
12 M. Cardona, R. V. Chamberlin, W. Marx, The history of the stretched exponential function. *Ann. Phys.* **16**, 842–845 (2007).
13 R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, I. C. Gunsalus, Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373 (1975).
14 A. T. N. Kumar, L. Zhu, J. F. Christian, A. A. Demidov, P. M. Champion, On the rate distribution analysis of kinetic data using the maximum entropy method: Applications to myoglobin relaxation on the nanosecond and femtosecond timescales. *J. Phys. Chem. B* **105**, 7847–7856 (2001).
15 J. Sabelko, J. Ervin, M. Gruebele, Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6031–6036 (1999).
16 J. F. Brandts, H. R. Halvorson, M. Brennan, Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry* **14**, 4953–4963 (1975).
17 D. Thirumalai, C. Hyeon, RNA and protein folding: Common themes and variations. *Biochemistry* **44**, 4957–4970 (2005).
18 G. R. Bowman, V. S. Pande, Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10890–10895 (2010).
19 B. Ahmad, Y. Chen, L. J. Lapidus, Aggregation of α-synuclein is kinetically controlled by intramolecular diffusion. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2336–2341 (2012).
20 P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
21 I. Peran *et al.*, Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12301–12310 (2019).