The Road Less Traveled in Protein Folding: Evidence for Multiple Pathways

Lisa J. Lapidus

Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824

lapidus@msu.edu

517-884-5656

Abstract

Free Energy Landscape theory of Protein Folding, introduced over 20 years ago, implies that a protein has many paths to the folded conformation with the lowest free energy. Despite the knowledge in principle, it has been remarkably hard to detect such pathways. The lack of such observations is primarily due to the fact that no one experimental technique can detect many parts of the protein simultaneously with the time resolution necessary to see such differences in paths. However, recent technical developments and employment of multiple experimental probes and folding prompts have illuminated multiple folding pathways in a number of proteins that had all previously been described with a single path.

Introduction

The past several decades have been filled with descriptions of the protein folding problem and potential solutions. More than one review paper over the past 2 decades has claimed that the problem is, more or less, solved [1]. Yet there is still no predictive theory to describe how a disordered chain of a specific peptide sequence (primary structure) folds to a stable 3-dimensional structure. And that simple proposition does not even cover the majority of biological examples discovered since the protein problem was first posed such as intrinsically disordered proteins that fold upon binding or never fold at all [2-4], co-translational folding on the ribosome [5,6], either with or without co-chaperones, refolding in large ATP-activated chaperone complexes after cellular stress [7-9], or folding in the crowded environment of the cell [10-12].

So what can account for the poor track record of the field? The most obvious one is that this is one of the most difficult problems in all of molecular biology or biophysics. The phase space that is occupied by an unfolded protein before folding is enormous and is difficult for scientists to comprehend, even computationally. The fact that macromolecular complexes million times larger than the smallest proteins can be more successfully simulated suggests that the constraints of structure are incredibly important. Furthermore, the smallest proteins are much larger than molecules that have been well studied spectroscopically and requires classical approximations of computational force fields. This puts the field of protein folding in the most difficult of halfway positions between the length and time scales that have been defined by chemistry and physics.

Nevertheless, we can count a number of achievements over the past few decades. Probably the first major hegemonic change after the problem was posed by Anfinson was the development of energy landscape theory, which posited that a multidimensional free energy surface described the path(s) that a protein traversed to the lowest free energy state [13,14]. That proteins did this in less than the age of the universe [15] meant that this landscape had to be minimally frustrated, in other words, shaped like a funnel. The implication of a multidimensional free energy landscape was that multiple pathways were possible and even likely, but this hypothesis has turned out to be remarkably difficult to demonstrate experimentally.

There are numerous reasons for this experimental bottleneck but they essentially come down to identifying the appropriate reaction coordinates and observing them with sufficient time resolution [16,17]. The implication of the minimally frustrated free energy landscape is that no matter which path a protein takes to folding, the time to traverse the landscape should be about the same. So kinetic signatures should look reasonably exponential. Additionally, the nature of most experimental probes is that they are generally local, such as the solvent exposure of a particular amino acid side chain. So subtle differences in kinetic signatures for different paths requires observing kinetics with multiple probes, an often difficult experimental challenge. Finally, another nature of the free energy funnel is that the diversity of paths will be most obvious further away from the folded state, which requires observing folding at the earliest time scales possible [18]. This precludes the best atomic-level probes, such as NMR, from providing much insight.

A further reason for the difficulty in observing multiple pathways in protein folding may not be the nature of protein folding, but the psychology of protein folding researchers. It is naturally easier to visualize, model and interpret data with a one-dimensional than a multi-dimensional free energy landscape. As researchers we have been admirably taught to follow Occam's razor and use the simplest explanation for our observations. Given the limited amount of information an experimentalist observing a folding protein has available, a 1-d interpretation is the natural outcome [19-21]. But you can also find similar approaches in computational studies where all atomic information is available [22-24]. This again supports the idea that scientists are drawn to simple explanations and also the extremely important goal of connecting simulation to experiment. Given these understandable limitations in our thinking, most conclusions of multiple pathways in protein folding have been made because no simpler explanation is available.

New methods to detect multiple pathways in bulk experiments

The most straightforward method to examine multiple pathways in protein folding is by making mutations to the sequence and comparing folding rates and equilibrium stabilities. However, these measurements don't usually reveal results that can unambiguously be interpreted as multiple pathways rather than perturbation of a single pathway. Indeed, the entire method of ϕ-value analysis implicitly assumes a single pathway. More generally, it is hard to know whether the mutations are creating new pathways or just changing the pathways already extant in the wildtype protein. Nevertheless, mutations remain a popular methodology for illuminating folding pathways.

A classic example of the evolution of thinking about protein folding pathways is the B1 domain of protein G. Protein G was one of the first small proteins to have its structure determined and researchers have been studying it ever since. This protein is very small and therefore conducive to MD simulation but has a complex fold with mixed secondary structure and is remarkably stable under a variety of conditions. The first structural measurements in equilibrium indicated the protein was "two-state" [25], that is, it only existed in the completely unfolded or completely folded state and that the entire structure formed cooperatively. The earliest kinetic folding studies, dilution of chemical denaturant using a stopped-flow mixer, confirmed the two state hypothesis [26,27]. However, the invention of more rapid mixing techniques suggested additional kinetic phases which would negate the two-state model, though not necessarily multiple pathways [28-30]. This conclusion was not uncontroversial [31]. Simultaneously, computational studies found both agreement and disagreement with the two-state model [32-34].

Most of these initial studies either used NMR, which did not have the time resolution of to look for sub-millisecond kinetic phases, or tryptophan fluorescence which only probes the local environment of one side chain. With only one time-resolved probe, it is unsurprising that the data did not demand more than one folding pathway. However, when my lab used multiple kinetic probes on the sub-millisecond time scale, multiple kinetic phases were observed. When Vijay Pande's group, then at Stanford University, constructed a Markov State Model from many

short simulations, multiple pathways were found and the reconstruction of the kinetic observables gave both qualitative and quantitative agreement [35].

More recent experimental work on using multiple probes of folding have also yielded multiple folding pathways. Charlier et al. developed a pressure jump (increase or decrease) instrument coupled to an NMR spectrometer, that allows time-resolved observation of individual resonances. Peaks identified as an on-pathway intermediate were observed, but they never dominate the entire population; the majority of the population proceeds via a 2-state folding mechanism [36]. Similarly, Nagarajan used site-specific labeling of different residues of the villin headpiece HP-36 for detection by infrared spectroscopy and found different kinetics at different locations [37].

Another way for multiple pathways to become apparent is with different prompts of folding, such as pressure, temperature or denaturant concentration. Wirth et al. have developed a pressure drop (decrease only) instrument with even higher time resolution than Charlier to study the microsecond folding the FiP35 WW domain [38]. When compared with T-jump experiments, this study finds distinctly different kinetic phases when folding is observed by the different methods, leading to an interpretation of two parallel pathways in which each of the two hairpins forms first. These results are qualitatively similar to an earlier studies on λ-repressor and villin headpiece HP-35 comparing T-jump to rapid mixing to dilute denaturant [39,40]. In both the FiP35 and λ-repressor study, T-jump kinetics were faster than either P-drop of denaturant dilution. This suggests the rapid increase in temperature by ~10 C near the temperature midpoint does not unfold the protein in same way as either high pressure or high denaturant. If multiple pathways are high populated over the entire free energy landscape, perhaps T-jump does not sample all these conformations as fully as other denaturation methods [41]. However, multiple pathways were only inferred by comparing different folding prompts.

The most common way for multiple pathways to be resolved is by all atom simulation, provided enough simulation time of the unfolded state allows the protein to explore enough phase space to find multiple pathways. Recently, significant effort has been made to recreate experimental observables to illustrate how the multiple pathways. Zanetti-Polzi et al. simulated the FiP35, found multiple pathways within the simulated folding trajectories, calculated IR absorption spectra for each path and found different IR signatures, but one of the intermediates has a very similar spectrum to the folded state and was therefore not detected independently by experiments [42]. A later study with the same collaboration looked at simulations and experiments of a different WW domain, GTT35, was able to observe different kinetics for different IR frequencies experimentally as well as computationally [43].

Altering pathways in the cellular context

Another approach that has revealed different folding pathways is part of the larger effort in the field to compare *in vitro* and *in vivo* folding. The best efforts to date involve co-translational folding of a nascent chain attached to a ribosome. Marsden et al. compared folding rates of circular permutants while attached to the ribosome to determine which secondary structure

motifs fold first but ultimately find a single pathway is sufficient to explain the data [44]. Samelson et al measured the folding of HaloTag on and off the ribosome and found that co-translational folding did not populate an intermediate that leads to aggregation [45]. Similarly Liu et al, used single molecule force spectroscopy to measure folding and unfolding of multi-domain elongation-factor G (EF-G) and found both the ribosome and trigger factor made folding more efficient, but ultimately found hierarchical folding of the domains as they emerge from the tunnel [46]. Guinn et al., using similar techniques, found src SH3 followed the same pathway on and off the ribosome [47]. However, in another paper, Guinn et al could find a separate pathway if the force was applied at different attachment points [48].

Single Molecule Experiments

A different approach to looking for multiple pathways experimentally is to use single molecule methods. Since individual proteins are probed separately, it should be possible to distinguish whether a particular protein or trajectory of a protein probed multiple times, follows one pathway or another. In reality, this is difficult to do because of intrinsic noise of individual measurements which in turn limits the time resolution or forces the experimenter to combine data from many measurements or apply single pathway models to the data. Another limitation of single molecule experiments is the need to work under conditions in which multiple folding and unfolding events can be reported for the same protein in order to collect enough data. This results in measurements far from native conditions (e.g. using some denaturant or applying a force), which may smooth out the landscape or eliminate one or more pathways.

For example, my lab, in collaboration with Matt Comstock and Michigan State University, measured folding and unfolding kinetics of individual molecules or protein G using laser tweezers. Previous work by Lee et al. showed Arrhenius kinetics of unfolding with force such that the unfolding rate increased with force, but the nature of their instrument required high forces (> 20 pN) in which the protein was only unfolded as the force was increased and each protein was cycled many times between folded and unfolded. A more sensitive instrument by our groups allow observation of folding and unfolding events near native conditions. We observed a turnover in unfolding (but not folding) times at forces near 5 pN such that the unfolding rate decreased with force at low forces [49]. Nevertheless, these measurements required multiple events on multiple molecules to see this trend. The second pathway that emerged at low forces, was only evident because the distinctive turnover (see Figure 2).

Most single molecule experiments using fluorescence resonance energy transfer or molecules tethered to a surface have resolved only 2-state behavior or obligate intermediates that could be justified by a single pathway. But groups can find multiple pathways if they employ methods to look for them. Hidden Markov modeling (HMM) has been used, most thoroughly, with adenosine kinase to delineate multiple pathways [50]. This analysis requires a designation of a certain number of states with different experimental observables, such as FRET levels, and analyzes the trajectories to find conditional probabilities of occupying each state based on the previous state. Aviram et al. applied this method to protein L, a classic two state folder that has

nevertheless been described with a complex free energy landscape for early folding events, and found a second pathway involving an intermediate that is populated at less than 10%. With the assumption of 6 states for adenosine kinase, Kantaev et al, found 7 pathways utilizing different combinations of these states. The sequential pathway, visiting all 6 states in order was found to be the most likely path and is dominant in higher denaturant concentrations, but more complexity could be found at low denaturant, making the cautionary point that denaturing conditions may smooth out free energy landscapes. Kantaev et al also found that a destabilizing mutant accessed more non-sequential pathways, suggesting that mutation may be provide a route to finding less populated paths [51].

One of the best cases in which single molecule experiments could directly reveal complex pathways was found for SOD1. Sen Mojumadar et al used single molecule force experiments to find 7 intermediate states, as revealed by jumps between distinct contour lengths representing the beta strands that unfolded. Bulk experiments had generally revealed only a two state model. As the protein folds or refolds, different molecules (or different pulls on the same molecule) sampled different intermediate states. The distinct intermediates allowed the logging of pairwise transitions to map multiple pathways and found most were not obligate [52]. While this work represents the most detailed account of multiple folding pathways from a single molecule experiment, interpretation was only possible by the fortuitous circumstance of separately resolvable intermediates.

Conclusions and Outlook

While both the intentional search for and fortuitous finding of multiple pathways have been features of some recent protein folding experiments, this is not the most efficient way forward. One recent advance is the development of analysis tools, such as hidden Markov methods, that can be applied to experimental data. Tavakoli et al. employed Bayesian nonparametric methods to determine the number of molecules in a single molecule experiment of diffusing proteins directly from photon arrival times, thus eliminating the need to combine many molecules in fluorescence correlation spectroscopy . Matsunaga and Sugita have recently used machine learning to create a Markov state model base on both molecular dynamics simulations and FRET data by Chung et al on the FBP WW domain to delineate multiple pathways [53,54]. At the other end of the complexity spectrum, Gopi et al used the Wako-Saitô-Muñoz-Eaton, a structure based model, to define all the microstates of a protein and used Monte Carlo simulations to follow pathways from each microstate to all others [55]. Applying this method to 5 small proteins typically observed to be two-state, WW domain, gpW, SH3, PDD and CspB, they find all of them have multiple pathways with significant flux.

In conclusion, protein folding researchers now have a good number of tools in their toolbox to look for evidence of multiple pathways in protein folding. However, it will never be as straightforward to delineate multiple paths rather than a single one. Thus, the incentive to find such pathways must be there for scientists to do the harder work. One possible incentive is the ability to understand how cellular environments alter pathway probabilities, the propensity for

aggregation and how interactions with chaperones prevent such aggregation; in other words, how does folding interact with the rest of cellular processes? This is surely a road worth exploring.

The author declares no conflict of interest

References

1. Wolynes PG, Eaton WA, Fersht AR: **Chemical physics of protein folding**. *Proceedings of the National Academy of Sciences* 2012, **109**:17770-17771.
2. Bonetti D, Troilo F, Brunori M, Longhi S, Gianni S: **How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein?** *Biophysical Journal* 2018, **114**:1889-1894.
3. Chong S-H, Ham S: **Folding Free Energy Landscape of Ordered and Intrinsically Disordered Proteins**. *Scientific Reports* 2019, **9**.
4. Toto A, Malagrino F, Visconti L, Troilo F, Pagano L, Brunori M, Jemth P, Gianni S: **Templated folding of intrinsically disordered proteins**. *The Journal of biological chemistry* 2020, **295**:6586-6593.
5. Kaiser CM, Liu K: **Folding up and Moving on-Nascent Protein Folding on the Ribosome**. *Journal of Molecular Biology* 2018, **430**:4580-4591.
6. Liutkute M, Samatova E, Rodnina MV: **Cotranslational Folding of Proteins on the Ribosome**. *Biomolecules* 2020, **10**.
7. Balchin D, Hayer-Hartl M, Hartl FU: **In vivo aspects of protein folding and quality control**. *Science* 2016, **353**.
8. Balchin D, Milicic G, Strauss M, Hayer-Hartl M, Hartl FU: **Pathway of Actin Folding Directed by the Eukaryotic Chaperonin TRiC**. *Cell* 2018, **174**:1507-+.
9. Bhatt JM, Enriquez AS, Wang J, Rojo HM, Molugu SK, Hildenbrand ZL, Bernal RA: **Single-Ring Intermediates Are Essential for Some Chaperonins**. *Frontiers in Molecular Biosciences* 2018, **5**.
10. Cheng C, Wu J, Liu G, Shi S, Chen T: **Effects of Non-native Interactions on Frustrated Proteins Folding under Confinement**. *Journal of Physical Chemistry B* 2018, **122**:7654-7667.
11. Sorokina I, Mushegian A: **Modeling protein folding in vivo**. *Biology Direct* 2018, **13**.
12. Wang Y, Sukenik S, Davis CM, Gruebele M: **Cell Volume Controls Protein Stability and Compactness of the Unfolded State**. *The Journal of Physical Chemistry B* 2018, **122**:11762-11770.
13. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG: **Funnels, Pathways, And The Energy Landscape Of Protein-Folding - A Synthesis**. *Proteins-Structure Function And Genetics* 1995, **21**:167-195.
14. Onuchic JN, LutheySchulten Z, Wolynes PG: **Theory of protein folding: The energy landscape perspective**. *Annual Review of Physical Chemistry* 1997, **48**:545-600.
15. Levinthal C: **How to fold graciously.** In *Mossbauer Spectroscopy in Biological Systems*; *Monticello, Illinois*,Edited by De Brunner P, Tsibris J, Munck E: University of Illinois Press: 1969:22-24.
16. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J: **Fast kinetics and mechanisms in protein folding**. *Annual Review of Biophysics and Biomolecular Structure* 2000, **29**:327-359.
17. Chong S-H, Ham S: **Examining a Thermodynamic Order Parameter of Protein Folding**. *Scientific Reports* 2018, **8**.
18. Lapidus LJ: **Exploring the top of the protein folding funnel by experiment**. *Current Opinion in Structural Biology* 2013, **23**:30-35.

19. Bergasa-Caceres F, Rabitz HA: **Predicting the location of the non-local contacts in alpha-synuclein**. *Biochimica Et Biophysica Acta-Proteins and Proteomics* 2018, **1866**:1201-1208.
20. da Silva FB, Contessoto VG, de Oliveira VM, Clarke J, Leite VBP: **Non-Native Cooperative Interactions Modulate Protein Folding Rates**. *Journal of Physical Chemistry B* 2018, **122**:10817-10824.
21. Troilo F, Malagrino F, Visconti L, Toto A, Gianni S: **The Effect of Proline cis-trans Isomerization on the Folding of the C-Terminal SH2 Domain from p85**. *International Journal of Molecular Sciences* 2020, **21**.
22. Maity H, Reddy G: **Transient intermediates are populated in the folding pathways of single-domain two-state folding protein L**. *Journal of Chemical Physics* 2018, **148**.
23. Shao Q, Yang L, Zhu W: **Selective enhanced sampling in dihedral energy facilitates overcoming the dihedral energy increase in protein folding and accelerates the searching for protein native structure**. *Physical Chemistry Chemical Physics* 2019, **21**:10423-10435.
24. Shao Q, Zhu W: **Nonnative contact effects in protein folding**. *Physical Chemistry Chemical Physics* 2019, **21**:11924-11936.
25. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM: **A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G**. *Science* 1991, **253**:657-661.
26. Kuszewski J, Clore GM, Gronenborn AM: **Fast folding of a prototypic polypeptide: The immunoglobulin binding domain of streptococcal protein G**. *Protein Science* 1994, **3**:1945-1952.
27. McCallister EL, Alm E, Baker D: **Critical role of beta-hairpin formation in protein G folding**. *Nat Struct Biol* 2000, **7**:669-673.
28. Park SH, ONeil KT, Roder H: **An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core**. *Biochemistry* 1997, **36**:14277-14283.
29. Park SH, Shastry MCR, Roder H: **Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing**. *Nature Structural Biology* 1999, **6**:943-947.
30. Park S-H, O'Neil KT, Roder H: **An Early Intermediate in the Folding Reaction of the B1 Domain of Protein G Contains a Native-like Core**. *Biochemistry* 1997, **36**:14277-14283.
31. Krantz BA, Mayne L, Rumbley J, Englander SW, Sosnick TR: **Fast and Slow Intermediate Accumulation and the Initial Barrier Mechanism in Protein Folding**. *Journal of Molecular Biology* 2002, **324**:359-371.
32. Cheng Q, Joung I, Lee J, Kuwajima K, Lee J: **Exploring the Folding Mechanism of Small Proteins GB1 and LB1**. *Journal of Chemical Theory and Computation* 2019, **15**:3432-3449.
33. Zagrovic B, Sorin E, Pande V: **Distributed computing simulations of folding of beta hairpin from protein G and villin headpiece**. *Biophysical Journal* 2002, **82**:474a-474a.
34. Brown S, Head-Gordon T: **Intermediates and the folding of proteins L and G**. *Protein Science* 2004, **13**:958-970.
35. Lapidus LJ, Acharya S, Schwantes CR, Wu L, Shukla D, King M, DeCamp SJ, Pande VS: **Complex Pathways in Folding of Protein G Explored by Simulation and Experiment**. *Biophysical Journal* 2014, **107**:947-955.

36. Charlier C, Alderson TR, Courtney JM, Ying J, Anfinrud P, Bax A: **Study of protein folding under native conditions by rapidly switching the hydrostatic pressure inside an NMR sample cell**. *Proceedings of the National Academy of Sciences of the United States of America* 2018, **115**:E4169-E4178.
37. Nagarajan S, Xiao S, Raleigh DP, Dyer RB: **Heterogeneity in the Folding of Villin Headpiece Subdomain HP36**. *Journal of Physical Chemistry B* 2018, **122**:11640-11648.
38. Wirth AJ, Liu Y, Prigozhin MB, Schulten K, Gruebele M: **Comparing Fast Pressure Jump and Temperature Jump Protein Folding Experiments and Simulations**. *Journal of the American Chemical Society* 2015, **137**:7152-7159.
39. DeCamp SJ, Naganathan AN, Waldauer SA, Bakajin O, Lapidus LJ: **Direct Observation of Downhill Folding of lambda-Repressor in a Microfluidic Mixer**. *Biophysical Journal* 2009, **97**:1772-1777.
40. Zhu L, Ghosh K, King M, Cellmer T, Bakajin O, Lapidus LJ: **Evidence of Multiple Folding Pathways for the Villin Headpiece Subdomain**. *The Journal of Physical Chemistry B* 2011, **115**:12632-12637.
41. Ghosh K, Ozkan SB, Dill KA: **The Ultimate Speed Limit to Protein Folding Is Conformational Searching**. *Journal of the American Chemical Society* 2007, **129**:11920-11927.
42. Zanetti-Polzi L, Davis CM, Gruebele M, Dyer RB, Amadei A, Daidone I: **Parallel folding pathways of Fip35 WW domain explained by infrared spectra and their computer simulation**. *FEBS Letters* 2017, **591**:3265-3275.
43. Davis CM, Zanetti-Polzi L, Gruebele M, Amadei A, Dyer RB, Daidone I: **A quantitative connection of experimental and simulated folding landscapes by vibrational spectroscopy**. *Chemical Science* 2018, **9**:9002-9011.
44. Marsden AP, Hollins JJ, O'Neill C, Ryzhov P, Higson S, Mendonca CATF, Kwan TO, Kwa LG, Steward A, Clarke J: **Investigating the Effect of Chain Connectivity on the Folding of a Beta-Sheet Protein On and Off the Ribosome**. *Journal of Molecular Biology* 2018, **430**:5207-5216.
45. Samelson AJ, Bolin E, Costello SM, Sharma AK, O'Brien EP, Marqusee S: **Kinetic and structural comparison of a protein's cotranslational folding and refolding pathways**. *Science Advances* 2018, **4**.
46. Liu K, Maciuba K, Kaiser CM: **The Ribosome Cooperates with a Chaperone to Guide Multi-domain Protein Folding**. *Molecular Cell* 2019, **74**:310-+.
47. Guinn EJ, Tian P, Shin M, Best RB, Marqusee S: **A small single-domain protein folds through the same pathway on and off the ribosome**. *Proceedings of the National Academy of Sciences of the United States of America* 2018, **115**:12206-12211.
48. Guinn EJ, Jagannathan B, Marqusee S: **Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein**. *Nature Communications* 2015, **6**:6861.
49. Izadi D, Chen Y, Whitmore ML, Slivka JD, Ching K, Lapidus LJ, Comstock MJ: **Combined Force Ramp and Equilibrium High-Resolution Investigations Reveal Multipath Heterogeneous Unfolding of Protein G**. *The Journal of Physical Chemistry B* 2018, **122**:11155-11165.
50. Taylor JN, Pirchi M, Haran G, Komatsuzaki T: **Deciphering hierarchical features in the energy landscape of adenylate kinase folding/unfolding**. *The Journal of Chemical Physics* 2018, **148**:123325.

51. Kantaev R, Riven I, Goldenzweig A, Barak Y, Dym O, Peleg Y, Albeck S, Fleishman SJ, Haran G: **Manipulating the Folding Landscape of a Multidomain Protein**. *Journal of Physical Chemistry B* 2018, **122**:11030-11038.
52. Sen Mojumdar S, N. Scholl Z, Dee DR, Rouleau L, Anand U, Garen C, Woodside MT: **Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories**. *Nature Communications* 2017, **8**:1881.
53. Matsunaga Y, Sugita Y: **Linking time-series of single-molecule experiments with molecular dynamics simulations by machine learning**. *Elife* 2018, **7**.
54. Chung HS, McHale K, Louis JM, Eaton WA: **Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times**. *Science* 2012, **335**:981-984.
55. Gopi S, Singh A, Suresh S, Paul S, Ranu S, Naganathan AN: **Toward a quantitative description of microscopic pathway heterogeneity in protein folding**. *Physical Chemistry Chemical Physics* 2017, **19**:20891-20903.

Annotated References

Charlier et al.

** A new method of rapidly (within 3 ms) raise and lower the pressure within a NMR sample tube to observe folding and unfolding. Bidirectional switching and use of a global probe such as NMR, make this technique much more generally useful than previous methods of pressure denaturation.

Zanetti-Polzi et al.

** The coupling of simulation and experiment allowed the delineation and confirmation of mutple pathways. The authors take the important step in directly calculating experimental observables to show that one pathway was not distinguishable from the native state in the experiment.

Samelson et al.

* The multiple probes in this study allowed the comparison of folding of a protein on and off the ribosome. When folding co-translationally, an intermediate that may be an aggregation precursor was repressed.

Kantaev et al.

** A tour-de-force in modeling of single molecule FRET measurements of folding. Hidden Markov modeling found five different states and seven different pathways from the data. Mutations and denaturant changed the partitioning of these pathways.

Liu et al.

* One of the most complete descriptions of co-translational protein folding using single-molecule force spectroscopy. The interactions of the protein with the ribosome and trigger factor are separately delineated.

Sen Mojumdar et al.

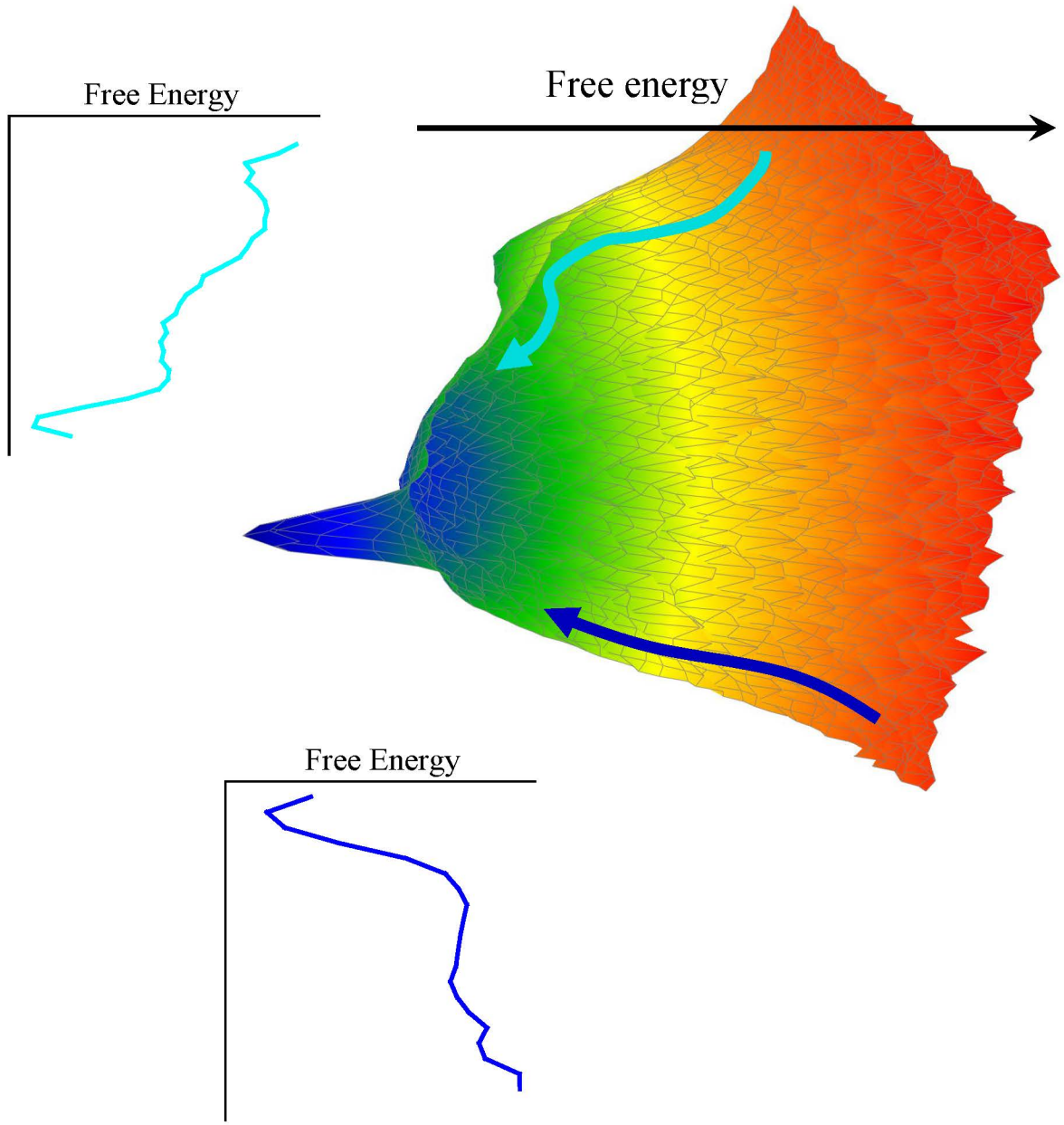
** The best direct experimental evidence of multiple folding pathways determined from single molecule force spectroscopy. As the protein is pulled and relaxed, individual

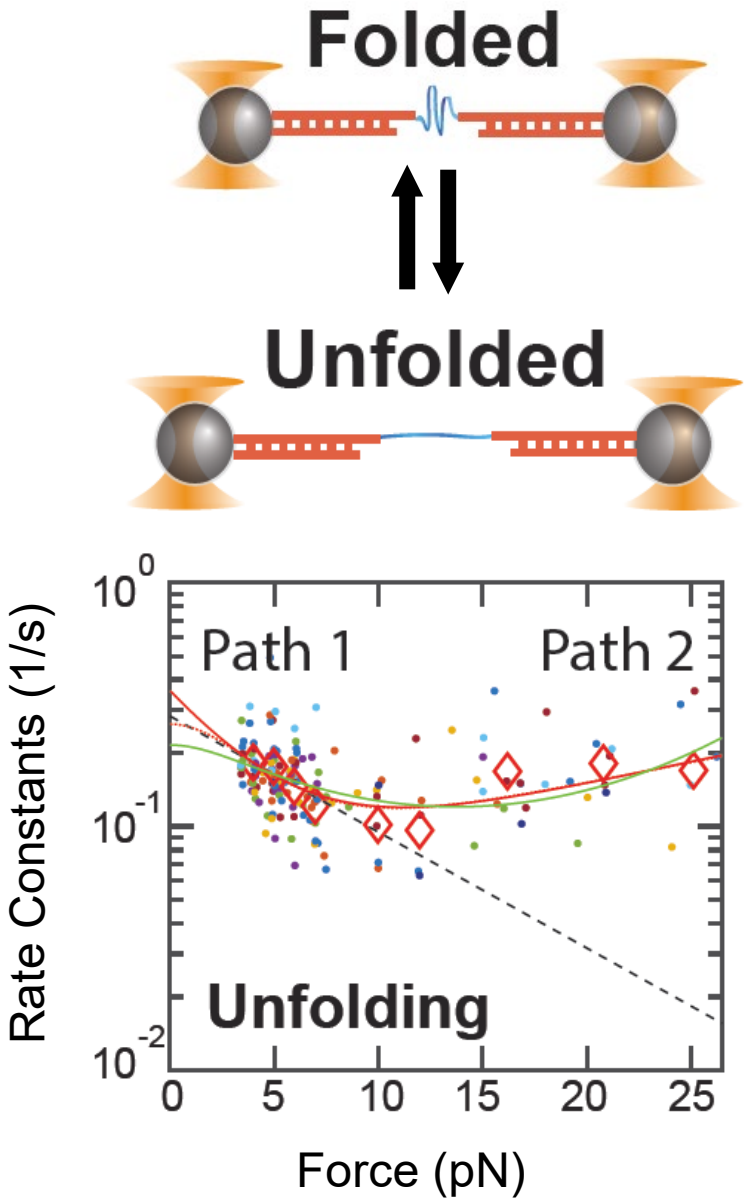
intermediates were detected from extension of secondary structure. These states were individually accessible in different pathways. This protein had generally been thought to be two-state prior to this study.

Figure Captions

Figure 1. Protein Folding free energy landscape. The lowest point (shaded blue) is the folded state, which can be approached via different pathways (blue and cyan arrows). Depending on which pathway is probed by a particular experiment, the interpretation of folding may be either 2-state (blue) or 3-state (cyan).

Figure 2. Example of 2 unfolding pathways observed by single molecule methods. (top) The B1 domain of protein G is connected between two long DNA handles which are connected to micron-sized beads held in separate optical traps. When on one trap is displaced relative to the other, a force is exerted on the protein, causing it to unfold. Holding the displacement constant allows for observation of multiple folding and unfolding events of a single molecule at different forces. (bottom) Unfolding rate constants extracted from the constant displacement measurements at various forces. Each point represents a separate molecule, the red diamonds are the weighted average of all molecules at a particular force. The solid lines are two separate models involving 2 independent unfolding pathways. The dashed line is the expected rates from just the lowest force measurements.


Free Energy
Free energy
Free Energy


Folded
Unfolded
Path 1
Path 2
Unfolding
Rate Constants (1/s)
Force (pN)
10^0
10^-1
10^-2
0
5
10
15
20
25