Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

# Understanding the heterogeneity intrinsic to protein folding

Sandhya Bhatia[1] and Jayant B. Udgaonkar[2]

**Abstract**

Relating the native fold of a protein to its amino acid sequence remains a fundamental problem in biology. While computer algorithms have demonstrated recently their prowess in predicting *what* structure a particular amino acid sequence will fold to, an understanding of *how* and *why* a specific protein fold is achieved remains elusive. A major challenge is to define the role of conformational heterogeneity during protein folding. Recent experimental studies, utilizing time-resolved FRET, hydrogen-exchange coupled to mass spectrometry, and single-molecule force spectroscopy, often in conjunction with simulation, have begun to reveal how conformational heterogeneity evolves during folding, and whether an intermediate ensemble of defined free energy consists of different sub-populations of molecules that may differ significantly in conformation, energy and entropy.

**Addresses**

[1] Department of Biophysics, Howard Hughes Medical Institute UT Southwestern Medical Center, Dallas 75390, United States

[2] Department of Biology, Indian Institute of Science Education and Research Pune, Pashan, Pune 41008, India

Corresponding author: Udgaonkar, Jayant B (jayant@iiserpune.ac.in) (Bhatia S.)

**Current Opinion in Structural Biology** 2024, **84**:102738

This review comes from a themed issue on **Folding and Binding (2024)**

Edited by **H. Jane Dyson** and **Peter E. Wright**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2023.102738



**Keywords**

Protein folding, Conformational heterogeneity, Folding pathways, Intermediate ensemble, Time-resolved FRET.

'*The ultimate rationale behind all purposeful structures and behaviour of living beings is embodied in the sequence of residues of nascent polypeptide chains — the precursors of the folded proteins which in biology play the role of Maxwell's demons. In a very real sense it is at this level of organization that the secret of life (if there is one) is to be found. If we could not only determine these sequences but also pronounce the law by which they fold, then the secret of life would be found — the ultimate rationale discovered!*' Jacques Monod [1,2].

Indeed, the fold of a protein determines its function, and the specific folds adopted by different protein sequences enable the multitude of chemical reactions and physical interactions that define life. How the one-dimensional amino acid sequence of a protein determines its functional three-dimensional native fold remains a fundamental problem in biology. One facet of the problem is to predict the functional structure from sequence, and here, recent progress has been remarkable [3]. Machine learning algorithms exploiting information so far hidden in the vast database of known protein structures, determined experimentally by scientists, can now predict protein folds with unprecedented accuracy [4,5]. Nevertheless, these algorithms have yet to lead to an understanding of the second facet of the folding problem which concerns how the protein fold is achieved [6].

Despite significant progress made towards understanding the mechanism of protein folding, many fundamental questions persist [7–10]. Perhaps the most fundamental question concerns how many folding pathways there are for a given protein. Examination of protein structures suggests that the final structure could assemble in many different ways, each defined by a specific order of conformational transitions and involving structurally distinct intermediates. Such a possibility had been highlighted in an influential paper, well before multiple folding pathways were detected for any protein [11]. Understanding the different ways by which a protein can fold is critical for understanding of the relative importance of the different physico-chemical forces known to stabilize a fully folded structure [12], as well as to drive the sequence of structural events that define any given folding pathway. Unfortunately, the ability to predict structure so much better has yet to lead to a better understanding of how these forces drive the folding process, although it seems likely that the machine learning algorithms must somehow have taken them into account, in order to be so successful.

There has long been awareness that the native state of a protein is structurally heterogeneous, that proteins exist "as a group of structures not too different from one another in free energy, but frequently differing considerably in energy and entropy. In fact, the molecule must be conceived as trying out every possible structure each in accordance with its Boltzmann factor" [13]. This conclusion was based on the results of hydrogen exchange studies; more recently, ion mobility spectrometry-mass spectrometry studies [14] have been able to detect a multiplicity of native states with extraordinary sensitivity. While the importance of structural heterogeneity and dynamics in facilitating protein function is becoming increasingly evident [15,16], much is yet to be known about the role of structural heterogeneity during the actual folding process. Of particular importance is the question of whether a folding intermediate ensemble consists of sub-populations of molecules, which are distinct in conformation but similar in free energy. How heterogeneity evolves on, and modulates, folding pathways is yet to be understood. Machine learning algorithms are already being used to explore protein folding pathways [17].

Delineation of the heterogeneity characteristic of protein folding pathways has important implications for understanding the mechanism of protein folding. It means that it cannot always be predicted how a protein will fold because structurally distinct sub-populations of an intermediate ensemble may be populated under different folding conditions. Hence, even more care has to be taken to interpret experimental data on protein folding. New experimental techniques and theoretical models have to be developed to better understand the folding pathways of proteins. This opinion is on the current understanding, derived largely from experimental studies, of protein folding heterogeneity.

## Heterogeneity seen through the eyes of experiment and theory

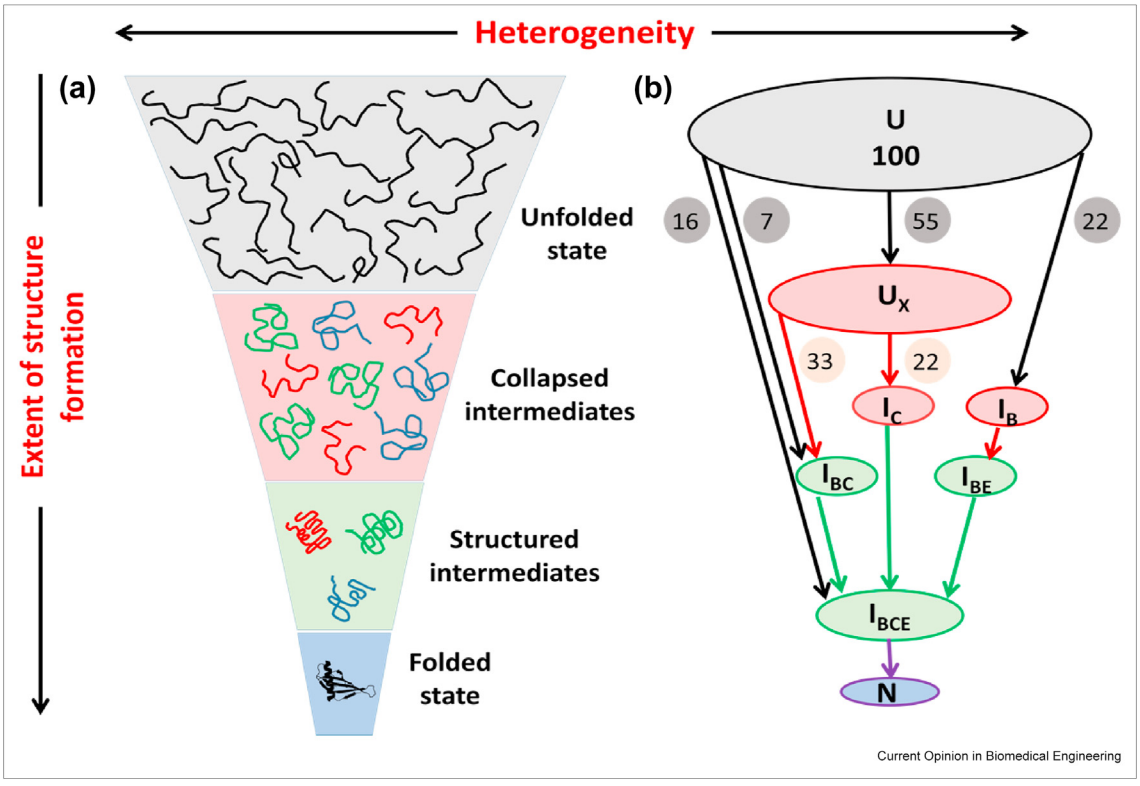
Protein folding occurs by the diffusive motion of the polypeptide chain starting from a highly dynamic U state [8,18,19]. The folding process is therefore expected to be heterogeneous. The idea that heterogeneity exists on protein folding pathways was legitimized when transiently populated partially folded intermediates on the pathways were characterized by hydrogen-exchange-NMR methods [20,21]. Subsequently, intermediates have been detected in the folding and unfolding pathways of many proteins. The initial thinking of many experimental investigators was that an intermediate constituted a relatively homogeneous population of partially folded molecules all with very similar structures. When the ideas of statistical mechanics and energy landscape views of folding became more prevalent [7,8,22], it became more accepted that all states on a protein folding pathway were ensembles of different conformations (Figure 1a). It was only through the concurrent use of multiple ensemble-averaging probes, or after the use of population-sensitive probes [23,24], that the underlying heterogeneity in folding intermediates could be identified experimentally (Figure 1b) [10,25]. Energy landscape theory, which could accommodate the results of many experimental studies, questioned whether folding intermediates have productive roles during folding, and also suggested that folding proceeded along a multiplicity of pathways on a rugged energy landscape [8,26,27]. On the other hand, experimental studies of folding kinetics, even those utilizing population-sensitive probes, have been unable to detect more than a few pathways for any given protein. Unfortunately, energy landscape theory cannot predict the structural mechanisms of (un)folding and cannot describe the nature of the pathways and how they differ structurally from each other.

Evidence for structural heterogeneity in folding intermediates has existed for a long time [21,23]. In fact, early studies suggested that folding intermediates and transition states can be structurally heterogeneous, consisting of sub-populations that differed in structure but were similar in free energy (Figure 1a). Because a different sub-population could be stabilized under different folding conditions, the intermediate structure and hence the folding pathway might appear different under different folding conditions. More recently, ion-mobility spectrometry-mass spectrometry measurements have been useful in teasing apart the heterogeneity present not only in the native states of proteins but also in their thermal unfolding reactions [14]. Heterogeneity can be modulated by changing the protein sequence [28], and also by the presence of other co-solutes and proteins [29–31]. More recent evidence has come from the use of high-resolution experimental methodologies that can track the folding process in real-time, [25,32–35]. Methods that can distinguish between multiple conformations present at the same time have been particularly useful. For example, a recent multi-site time-resolved FRET (tr-FRET) study of the folding of single-chain monellin (MNEI) has shown that an early intermediate ensemble consists of four sub-populations of folding molecules, whose structures could be distinguished (Figure 1b) [35]. Theoretical studies have begun to describe quantitatively microscopic pathway heterogeneity on protein folding pathways [36,37].

## Single dominant pathway *versus* multiple pathways

Early experiments investigated the folding reaction only along one or two reaction coordinates, by utilizing only one or two ensemble-averaging experimental probes such as fluorescence or circular dichroism. Hence, they

**Figure 1**




Heterogeneity in protein folding: Phenomenological model describing the evolution of structural heterogeneity as folding progresses from the unfolded (U) to the native (N) state. The extent of structure formation increases down the y-axis. Conformational heterogeneity decreases during folding, as depicted in the width of the shown schematic. The differently colored polypeptide chains in panel A represent conformationally distinct sub-populations of molecules with differential extents of compaction or structure formation. Different colors in the funnel in panel (a) are representative of different energy levels (highest on the top), depicted as four classes of distinct conformational ensembles: U state (grey), Collapsed intermediate (orange), Partially structured intermediates (green) and the N state (blue). The scheme in panel (b) is adapted from Bhatia et al., 2021, JACS [35]. It provides a quantitative description of the observed folding kinetics of MNEI probed using multi-site trFRET. The black, red, green, and violet arrows represent the unobservable (over within 100 ms), fast, slow, and very slow kinetic phases of the folding reaction, respectively. $U_x$ represents the unfolded state in refolding conditions. The subscripts, B, C, and E on the different intermediates (I) denote the different regions that have become structured. The numbers in the circles denote the percentages of molecules following a given folding route.

invariably described folding as occurring along a single defined pathway populated by folding intermediates that could be, in many cases, too sparsely populated to be detected [20,38]. Indeed, the notion of a single dominant folding pathway allowed the adoption of the elegant method of φ-value analysis [39] by many scientists, enabling elucidation of the structures of transition states and intermediates. The equally elegant hydrogen exchange NMR (HX-NMR) and hydrogen exchange mass spectrometry (HX-MS) methodologies have also appeared to support the view that folding occurs *via* intermediates that are progressively more structured on a single sequential pathway [38,40,41]. The belief that folding occurs *via* a unique pathway is very strong. Kinetic data on the folding of some proteins, which had been accounted for previously on the basis of multiple folding pathways, have been reanalyzed based on alternative mechanisms defined either by more states or more steps that all feed into the single pathway leading to the N state [38,42]. Of course, the invocation of additional conformational states also means the invocation of additional heterogeneity.

Evidence obtained early on, which showed that proteins can use multiple pathways to fold, has been reviewed earlier [23], and that evidence will not be repeated here. Recent time-resolved SAXS measurements have shown that the folding of cytochrome *c* occurs *via* multiple pathways [32]. A recent microsecond folding study [43] showed that the folding and unfolding of the C-terminal domain of the mouse prion protein occurred on the same pathway at pH 7, but on different pathways at pH 4. Very unusually, the pathways utilized at pH 4 depended on the initial conditions, and unfolding pathways could be switched by changing the initial conditions. Recently, multiple transition pathways have also been detected at

the single molecule level using a novel experimental approach involving nanopores [33].

Examining folding at the single molecule level, whether by experiment or by simulation, continues to reveal heterogeneity in folding reactions. Single-molecule force spectroscopy (SMFS) measurements have revealed multiple pathways of folding and unfolding for many proteins [34,44,45]. Importantly, the solution conditions under which the SMFS measurements are carried out, matter for revealing the existing folding pathway heterogeneity [46]. A recent SMFS study on talin revealed a complex energy landscape with distinct rare conformations that have physiological relevance, challenging conventional notions of equilibrium dynamics [47]. Diverse transition paths have been shown to operate during a binding-induced folding reaction [48]. Single-molecule FRET (smFRET) studies of the folding of multi-domain proteins, utilizing either measurements of multiple intramolecular distances [49] or measurements of a single intramolecular distance analyzed using Hidden Markov Modelling [50] have also revealed heterogeneous folding. Analysis of transition paths obtained from smFRET measurements has also explicitly shown conformationally distinct folding trajectories adopted by different sub-populations of molecules [51,52].

Molecular dynamics simulations have continued to discover multiple folding pathways, shedding light on the structural distinction between different pathways [27,28]. Denaturants have been shown to alter the fluxes through the multiple folding pathways of PDZ domain [53]. Both simulations [54] and experiments [50] have revealed that adenylate kinase utilizes multiple pathways to fold. Simulations have proven to be especially valuable when combined with experimental investigations of folding reactions [55]. Importantly, although certain simulations may initially suggest a single pathway due to the chosen reaction coordinate, further analysis using Markov state models has revealed the existence of multiple pathways [56].

In this context, it is important to examine the results of pulsed-labeling HX-NMR and HX-MS experiments. In such experiments, when carefully designed, the structure at the individual amino acid residue level must be seen to form in a single kinetic phase, and not in multiple stages. However, in one of the earliest pulsed labeling HX-NMR studies [21], of the folding of ribonuclease A, structure at the single residue level was seen to form in two kinetic phases, even after accounting for proline isomerization. Although the data could be described by invoking two competing pathways, they were analyzed on the basis of the folding of a fraction of molecules on a single pathway being held back by the presence of what was later referred to by others as optional errors [38,42]. Indeed, a later native-state HX-NMR study of ribonuclease A clearly suggested the presence of two(un)folding pathways [57]. In recent HX-MS studies of the folding of maltose binding protein [40] ribonuclease H [58] and cytochrome *c* [59], multi-exponential evolution of structure acquisition has been observed, even after accounting for potential errors [60]. While folding could certainly occur predominantly along one pathway in the case of these proteins [38], it is pertinent to note that an early classic HX-MS study of the turkey ovomucoid third domain uncovered multiple unfolding pathways [61]. Native state HX-MS investigations on single-chain monellin (MNEI) [62] and a SH3 domain [30] have also indicated that unfolding can occur through independent and parallel pathways. A pulsed-labeling HX-MS study of the folding of double-chain monellin [63] has revealed how the sequence of structural events during folding in strongly stabilizing conditions is distinct from that in less stabilizing conditions. This result was consistent with an earlier study that had indicated, on the basis of the properties of the transition state, that folding switched pathways under the two folding conditions [64].

## Assembly of structure can occur in multiple ways under the same folding conditions

The assembly of structural parts during folding can potentially occur in many distinct ways, even under the same folding conditions. It was important to demonstrate this structural distinction because it would validate the existence of concurrently operating folding pathways. Recent multi-site time-resolved FRET (trFRET) studies of MNEI have achieved this [35]. Earlier studies have indicated that MNEI folds and unfolds *via* multiple intermediates on multiple pathways [65,66]. A more recent study has indicated that part of the complexity arises from proline isomerization, but that competing pathways persist even when the *cis* proline residues are all replaced by alanine [67]. The trFRET studies have revealed how competing pathways can be distinct structurally [35], established that structural parts can fold independently in different molecules through multiple steps (Figure 1b), and that the same structural change can happen fast on one pathway and slow on another pathway. Interestingly, the pathway-averaged sequence of structure formation suggested a hierarchical accumulation of structure, where local contacts precede nonlocal contacts and global structural evolution. Notably, the sequence of structure formation derived from trFRET measurements correlated with the previously determined sequence of structure dissolution in native conditions, measured by HX-MS [68]. Importantly, this study showed that the observation of hierarchical structure formation does not exclude the existence of parallel folding pathways, challenging the previously proposed single-defined pathway model.

## Gradual conformational change

Both trFRET and HX-MS studies of the folding/unfolding of MNEI, as well as HX-MS studies of the folding/unfolding of a SH3 domain not only indicated that structural change can occur differently on different pathways but that it can occur gradually under some conditions [10,62]. In the case of MNEI, the gradual unfolding and folding reactions on competing pathways could be described adequately by a Rouse-like chain model [69] or by a coarse-grained Markov model [66]. Importantly, in the case of MNEI, (un)folding in the absence of denaturant was completely gradual in nature but could be tuned to be partly cooperative in the presence of even low concentrations of denaturant [68]. This suggested that the cooperativity or lack of heterogeneity commonly seen in the folding reactions of many proteins at low denaturant concentration may merely be a manifestation of the modulation of the energy landscape by the denaturant. Evidence supporting the partly downhill folding of barnase has been obtained from a recent calorimetric force spectroscopy study [70].

## The role of heterogeneity in protein conformational change

A full understanding of how proteins fold, of the heterogeneity inherent in the folding process, and of the physical and chemical forces that govern folding, is critical for a proper understanding of a variety of protein conformational changes. Does the ribosome modulate co-translational folding [31] by binding to one of several sub-populations present in an intermediate ensemble? What is the role of conformational heterogeneity in determining how the same sequence may fold into two very different conformations, as happens in the case of fold-switching proteins [71], as well as in the case of binding-induced folding of intrinsically disordered proteins [29,72]? What is its role in domain swapping that drives the formation of dimers and larger oligomers [73]? Does conformational heterogeneity play a role in the divergence of folding pathways of a protein family over evolutionary time [74]? And what is its role in driving the conversion of $\alpha$-helix to $\beta$-sheet, which characterizes many protein misfolding and aggregation processes that lead to disease [75]? Clearly, understanding conformational heterogeneity on and off protein folding pathways is critical for understanding many biological processes.

## Conclusion

Many detailed studies of the kinetic mechanisms of protein folding have shown that a significant number of single-domain and multi-domain proteins utilize multiple pathways to fold (reviewed in references 10, 23). Validation of the existence of competing pathways, by showing how they differ in the manner structure progressively forms on them, is, however, still at its early stages. Much of the kinetic and structural heterogeneity would be present at the early stages of folding, before the formation of molten globule forms, and smFRET studies of folding kinetics will be better able to characterize this heterogeneity when their temporal resolution is improved, and when multiple FRET pairs are utilized, as they have been in trFRET studies. It is only now, with the advent of methodologies such as trFRET and HX-MS that can distinguish structurally between multiple conformations present together and which can describe their temporal evolution, that the structural and temporal heterogeneity intrinsic to protein folding reaction is being revealed.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## Acknowledgments

We thank past and present members of the JBU laboratory, as well as G. Krishnamoorthy, M.K. Mathew and D. Dhar, for discussions. SB is a recipient of JCC/HHMI postdoctoral research fellowship. JBU is a recipient of a JC Bose National Fellowship from the Government of India. Work in the JBU laboratory has been funded by the Tata Institute of Fundamental Research, the Indian Institute of Science Education and Research Pune, and the Science and Engineering Research Board, Government of India.

## References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Monod J: *Chance and necessity: an essay on the natural philosophy of modern biology (translated into English by wainhouse, A.).* New York: Knopf; 1971:95–96.
2. Taylor WR, May ACW, Brown NP, Aszódi A: **Protein structure: geometry, topology and classification**. *Rep Prog Phys* 2001, **64**:517–590.
3. Jumper J, Hassabis D: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.
4. Baek M, Baker D: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science* 2021, **373**:871–876.
5. Moore PB, Hendrickson WA, Henderson R, Brunger AT: **The protein-folding problem: not yet solved**. *Science* 2022, **375**: 507.
6. Chen SJ, Rose GD: **Opinion: protein folds vs. protein folding: differing questions, different challenges**. *Proc Natl Acad Sci USA* 2023, **120**, e2214423119.
7. Thirumalai D, O'Brien EP, Morrison G, Hyeon C: **Theoretical perspectives on protein folding**. *Annu Rev Biophys* 2010, **39**: 159–183.
8. * Nassar R, Dignon GL, Razban RM, Dill KA: **The protein folding problem: the role of theory**. *J Mol Biol* 2021, **433**, 167126.

The review outlines how the protein folding problem, originally concerned with the emergence of native protein structures from disorder, has been addressed through theory and experiments, revealing principles like solvation-based coding and funnel-shaped conformational ensembles, expanding our understanding of protein science in biology.

9. ** Rose GD: **Protein folding - seeing is deceiving**. *Protein Sci* 2021, **30**:1606–1616.
This Perspective challenges the conventional interpretation of protein folding by proposing that high-energy excluding interactions, such as steric clashes and unsatisfied hydrogen bond donors and acceptors, play a significant role in reducing the accessible conformational space during folding, shedding light on the role of hydrogen-bond satisfaction as a key parameter in protein folding models.

10. ** Bhatia S, Udgaonkar JB: **Heterogeneity in protein folding and unfolding reactions**. *Chem Rev* 2022, **122**:8911–8935.
This review discusses the conformational heterogeneity in protein structures during folding reactions measured using high-resolution structural probes. It reveals that various states in protein folding, including unfolded, intermediate, and native states, exhibit significant conformational diversity, stemming from physicochemical interactions and functional constraints, which in turn contribute to the existence of multiple folding pathways.

11. Harrison SC, Durbin R: **Is there a single pathway for the folding of a polypeptide chain?** *Proc Natl Acad Sci USA* 1985, **82**: 4028–4030.

12. Tsuboyama K, *et al.*: **Mega Scale experimental analysis of protein folding stability in biology and design**. *Nature* 2023, **620**:434.

13. Linderstrøm-Lang K, Schellman JA: **Protein structure and enzyme activity**. In *In the enzymes*. New York: Academic; 1959: 443–510.

14. Raab SA, *et al.*: **Evidence for many unique solution structures for Chymotrypsin Inhibitor 2: a thermodynamic perspective derived from vT-ESI-IMS-MS measurements**. *J Am Chem Soc* 2020, **142**:17372–17383.

15. Alderson TR, Kay LE: **NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function**. *Cell* 2021, **184**:577–595.

16. Schwartz SD: **Protein dynamics and enzymatic catalysis**. *J Phys Chem B* 2023, **127**:2649–2660.

17. Zhao K, Xia Y, Zhang F, Zhou X, Li SZ, Zhang G: **Protein structure and folding pathway prediction based on remote homologs recognition using PAthreader**. *Commun Biol* 2023, **6**:243.

18. Takahashi S, Yoshida A, Oikawa H Hypothesis: **Structural heterogeneity of the unfolded proteins originating from the coupling of the local clusters and the long-range distance distribution**. *Biophys. Rev.* 2018, **10**:363–373.

19. Pastore A, Temussi PA: **The protein unfolded state: one, No one and one hundred thousand**. *J Am Chem Soc* 2022, **144**: 22352–22357.

20. Udgaonkar JB, Baldwin RL: **NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A**. *Nature* 1988, **335**:694–699.

21. Udgaonkar JB, Baldwin RL: **Early folding intermediate of ribonuclease A**. *Proc Natl Acad Sci USA* 1990, **87**:8197––8201.

22. Wolynes P, Onuchic J, Thirumalai D: **Navigating the folding routes**. *Science (Washington, DC, U S)* 1995, **267**:1619–1620.

23. Udgaonkar JB: **Multiple routes and structural heterogeneity in protein folding**. *Annu Rev Biophys* 2008, **37**:489–510.

24. Englander SW, Mayne L, Kan Z-Y, Hu W: **Protein folding-how and why: by hydrogen exchange, fragment separation, and mass spectrometry**. *Annu Rev Biophys* 2016, **45**:135–152.

25. * Chan AM, Nijhawan AK, Hsu DJ, Leshchev D, Rimmerman D, Kosheleva I, Kohlstedt KL, Chen LX: **The role of transient intermediate structures in the unfolding of the trp-cage fast-folding protein: generating ensembles from time-resolved X-ray solution scattering with genetic algorithms**. *J Phys Chem Lett* 2023, **14**:1133–1139.
The study provides valuable insights into the fast-folding dynamics of the Trp-cage miniprotein, revealing the existence of a conformationally extended intermediate on a microsecond timescale through time-resolved X-ray solution scattering, and offering a benchmark for theoretical models of protein folding ensembles produced with molecular dynamics simulations.

26. Eaton WA, Wolynes PG: **Theory, simulations, and experiments show that proteins fold by multiple pathways**. *Proc Natl Acad Sci USA* 2017, **114**:E9759–E9760.

27. Maity H, Reddy G: **Thermodynamics and kinetics of single-chain monellin folding with structural insights into specific collapse in the denatured state ensemble**. *J Mol Biol* 2018, **430**:465–478.

28. ** Chang L, Perez A: **Deciphering the folding mechanism of proteins G and L and their mutants**. *J Am Chem Soc* 2022, **144**: 14668–14677.
A unified approach combining simulations and Bayesian inference is used to elucidate the atomistic details of protein folding mechanisms for proteins G and L and their mutants, providing insights into their folding pathways and intermediate structures, with agreement with experimental data.

29. Daniels KG, Tonthat NK, McClure DR, Chang YC, Liu X, Schumacher MA, Fierke CA, Schmidler SC, Oas TG: **Ligand concentration regulates the pathways of coupled protein folding and binding**. *J Am Chem Soc* 2014, **136**: 822–825.

30. Jethva PN, Udgaonkar JB: **The osmolyte TMAO modulates protein folding cooperativity by altering global protein stability**. *Biochemistry* 2018, **57**:5851–5863.

31. Ahn M, Włodarski T, Mitropoulou A, Chan SHS, Sidhu H, Plessa E, Becker TA, Budisa N, Waudby CA, Beckmann R, Cassaignau AME, Cabrita LD, Christodoulou J: **Modulating co-translational protein folding by rational design and ribosome engineering**. *Nat Commun* 2022, **13**:4243.

32. Kim TW, Lee SJ, Jo J, Kim JG, Ki H, Kim CW, Cho KH, Choi J, Lee JH, Wulff M, Rhee YM, Ihee H: **Protein folding from heterogeneous unfolded state revealed by time-resolved X-ray solution scattering**. *Proc Natl Acad Sci U S A* 2020, **117**: 14996–15005.

33. Liu SC, Ying YL, Li WH, Wan YJ, Long YT: **Snapshotting the transient conformations and tracing the multiple pathways of single peptide folding using a solid-state nanopore**. *Chem Sci* 2021, **12**:3282–3289.

34. * Petrosyan R, Narayan A, Woodside MT: **Single-molecule force spectroscopy of protein folding**. *J Mol Biol* 2021, **433**, 167207.
This study provides an overview of the significant insights gained into protein folding by employing single-molecule force spectroscopy (SMFS), demonstrating its capacity to elucidate energy landscapes, folding pathways, chaperone mechanisms, co-translational folding, and membrane protein folding.

35. ** Bhatia S, Krishnamoorthy G, Udgaonkar JB: **Mapping distinct sequences of structure formation differentiating multiple folding pathways of a small protein**. *J Am Chem Soc* 2021, **143**:1447–1457.
This study addresses the long-standing challenge of experimentally characterizing structurally distinct multiple folding pathways. By employing multisite time-resolved FRET methodology, the research elucidates distinct structural differences between these pathways, providing a comprehensive model that describes the independent acquisition of structure in various segments of the protein.

36. Gopi S, Singh A, Suresh S, Paul S, Ranu S, Naganathan AN: **Toward a quantitative description of microscopic pathway heterogeneity in protein folding**. *Phys Chem Chem Phys* 2017, **19**:20891–20903.

37. Gianni S, Freiberger MI, Jemth P, Ferreiro DU, Wolynes PG, Fuxreiter M: **Fuzziness and frustration in the energy landscape of protein folding, function, and assembly**. *Acc Chem Res* 2021, **54**:1251–1259.

38. Englander SW, Mayne L: **The case for defined protein folding pathways**. *Proc Natl Acad Sci U S A* 2017, **114**:8253–8258.

39. Fersht AR, Itzhaki LS, elMasry NF, Matthews JM, Otzen DE: **Single versus parallel pathways of protein folding and**

**fractional formation of structure in the transition state**. *Proc Natl Acad Sci U S A* 1994, **91**:10426–10429.

40. Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW: **Folding of a large protein at high structural resolution**. *Proc Natl Acad Sci U S A* 2013, **110**:18898–18903.

41. Aghera N, Udgaonkar JB: **Stepwise assembly of β-sheet structure during the folding of an SH3 domain revealed by a pulsed hydrogen exchange mass spectrometry study**. *Biochemistry* 2017, **56**:3754–3769.

42. Bédard S, Krishna MMG, Mayne L, Englander SW: **Protein folding: independent unrelated pathways or predetermined pathway with optional errors**. *Proc Natl Acad Sci USA* 2008, **105**:7182–7187.

43. ** Moulick R, Goluguri RR, Udgaonkar JB: **Ruggedness in the free energy landscape dictates misfolding of the prion protein**. *J Mol Biol* 2019, **431**:807–824.
This study provides insights into the complex free energy landscape of the mouse prion protein, under different pH conditions, shedding light on how changes in initial conditions can alter folding and unfolding pathways.

44. Lapidus LJ: **The road less traveled in protein folding: evidence for multiple pathways**. *Curr Opin Struct Biol* 2021, **66**:83–88.

45. Neupane K, Woodside MT: **Hidden depths of protein folding**. *Nat Phys* 2023, **19**:8–9.

46. 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**. *J Phys Chem B* 2018, **122**:11155–11165.

47. ** Tapia-Rojo R, Mora M, Board S, Walker J, Boujemaa-Paterski R, Medalia O, Garcia-Manyes S: **Enhanced statistical sampling reveals microscopic complexity in the talin mechanosensor folding energy landscape**. *Nat Phys* 2023, **19**:52–60.
Single-molecule magnetic tweezers are used to explore protein folding and unfolding at extended equilibrium timescales, revealing previously unobserved complex conformational states in talin and shedding light on their physiological relevance, challenging the traditional understanding of equilibrium dynamics.

48. ** Kim J-Y, Chung HS: **Disordered proteins follow diverse transition paths as they fold and bind to a partner**. *Science* 2020.
Advanced single-molecule force spectroscopy reveals the diverse transition paths of a disordered protein, showcasing the significance of non-native interactions and chain flexibility in enabling faster binding compared to folded proteins.

49. Cerminara M, Schone A, Ritter I, Gabba M, Fitter J: **Mapping multiple distances in a multi-domain protein for the identification of folding intermediates**. *Biophys J* 2020, **118**:688–697.

50. 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**. *J Phys Chem B* 2018, **122**:11030–11038.

51. Eaton WA: **Modern kinetics and mechanism of protein folding: a retrospective**. *J Phys Chem B* 2021, **125**:3452–3467.

52. Makarov DE: **Barrier crossing dynamics from single-molecule measurements**. *J Phys Chem B* 2021, **125**:2467–2476.

53. Liu Z, Thirumalai D: **Denaturants alter the flux through multiple pathways in the folding of PDZ domain**. *J Phys Chem B* 2018, **122**:1408–1416.

54. * Liu Z, Thirumalai D: **Cooperativity and folding kinetics in a multidomain protein with interwoven chain topology**. *ACS Cent Sci* 2022, **8**:763–774.
A coarse-grained model is used to simulate the folding of a multiply connected multidomain protein (MMP), adenylate kinase (ADK), and provides insights into the cooperative interactions and interwoven chain topology that dictate its assembly mechanism.

55. 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 Lett* 2017, **591**:3265–3275.

56. Beauchamp KA, McGibbon R, Lin Y-S, Pande VS: **Simple few-state models reveal hidden complexity in protein folding**. *Proc Natl Acad Sci USA* 2012, **109**:17807–17813.

57. Juneja J, Udgaonkar JB: **Characterization of the unfolding of ribonuclease A by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding**. *Biochemistry* 2002, **41**:2641–2654.

58. Hu W, Walters BT, Kan ZY, Mayne L, Rosen LE, Marqusee S, Englander SW: **Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry**. *Proc Natl Acad Sci USA* 2013, **110**:7684–7689.

59. Hu W, Kan ZY, Mayne L, Englander SW: **Cytochrome c folds through foldon-dependent native-like intermediates in an ordered pathway**. *Proc Natl Acad Sci USA* 2016, **113**:3809–3814.

60. Englander SW, Mayne L, Kan Z-Y, Hu W: **Protein folding-how and why: by hydrogen exchange, fragment separation, and mass spectrometry**. *Annu Rev Biophys* 2016, **45**:135–152.

61. Arrington CB, Teesch LM, Robertson AD: **Defining protein ensembles with native-state NH exchange: kinetics of interconversion and cooperative units from combined NMR and MS analysis**. *J Mol Biol* 1999, **285**:1265–1275.

62. Malhotra P, Udgaonkar JB: **How cooperative are protein folding and unfolding transitions?** *Protein Sci* 2016, **25**:1924––1941.

63. * Bhattacharjee R, Udgaonkar JB: **Differentiating between the sequence of structural events on alternative pathways of folding of a heterodimeric protein**. *Protein Sci* 2022, **31**, e4513.
Pulsed hydrogen exchange labeling and mass spectrometry is employed to characterize the folding of double chain monellin (dcMN), revealing a triangular three-state folding mechanism that distinguishes between alternative pathways based on differences in the progress of structure acquisition, particularly in the β2 region, shedding light on the sequence of structural events in protein folding.

64. Aghera N, Udgaonkar JB: **Kinetic studies of the folding of heterodimeric monellin: evidence for switching between alternative parallel pathways**. *J Mol Biol* 2012 Jul 13, **420**:235–250.

65. Patra AK, Udgaonkar JB: **Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways**. *Biochemistry* 2007, **46**:11727–11743.

66. Bhatia S, Krishnamoorthy G, Dhar D, Udgaonkar JB: **Observation of continuous contraction and a metastable misfolded state during the collapse and folding of a small protein**. *J Mol Biol* 2019, **431**:3814–3826.

67. Kaushik A, Udgaonkar JB: **Replacement of the native cis prolines by alanine leads to simplification of the complex folding mechanism of a small globular protein**. *Biophys J* 2023, **3495**.00527-1.

68. Malhotra P, Udgaonkar JB: **Secondary structural change can occur diffusely and not modularly during protein folding and unfolding reactions**. *J Am Chem Soc* 2016, **138**:5866–5878.

69. Jha SK, Dhar D, Krishnamoorthy G, Udgaonkar JB: **Continuous dissolution of structure during the unfolding of a small protein**. *Proc Natl Acad Sci U S A* 2009, **106**:11113–11118.

70. Rico-Pasto M, Zaltron A, Davis SJ, Frutos S, Ritort F: **Molten globule-like transition state of protein barnase measured with calorimetric force spectroscopy**. *Proc Natl Acad Sci U S A* 2022, **119**, e2112382119.

71. Ruan B, He Y, Chen Y, Choi EJ, Chen Y, Motabar D, Solomon T, Simmerman R, Kauffman T, Gallagher DT, Orban J, Bryan PN: **Design and characterization of a protein fold switching network**. *Nat Commun* 2023, **14**:431.

72. Karlsson E, Andesson E, Camilloni C: **A structurally heterogeneous transition state underlies coupled binding and unfolding of disordered proteins**. *J Biol Chem* 2019, **294**: 1230–1239.

73. Nandwani N, Surana P, Negi H, Mascarenhas N, Udgaonkar JB, Das R, Gosavi S: **A five-residue motif for the design of domain swapping in proteins**. *Nat Commun* 2019, **10**:1–13.

74. Lim SA, Bolin ER, Marqusee S: **Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange**. *Elife* 2018, **7**, e38369.

75. Pal S, Udgaonkar JB: **Evolutionarily conserved proline residues impede the misfolding of the mouse prion protein by destabilizing an aggregation-competent partially unfolded form**. *J Mol Biol* 2022, **434**, 167854.