



Observation of Continuous Contraction and a Metastable Misfolded State during the Collapse and Folding of a Small Protein

Sandhya Bhatia¹, G. Krishnamoorthy², Deepak Dhar³ and Jayant B. Udgaonkar^{1,3}

¹ - National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560 065, India

² - Department of Biotechnology, Anna University, Chennai 600 025, India

³ - Indian Institute of Science Education and Research, Pune 411 008, India

Correspondence to G. Krishnamoorthy, Deepak Dhar and Jayant B. Udgaonkar: gk@tifr.res.in, deepak@iiserpune.ac.in, jayant@ncbs.res.in

<https://doi.org/10.1016/j.jmb.2019.07.024>

Edited by Sheena Radford

Abstract

To obtain proper insight into how structure develops during a protein folding reaction, it is necessary to understand the nature and mechanism of the polypeptide chain collapse reaction, which marks the initiation of folding. Here, the time-resolved fluorescence resonance energy transfer technique, in which the decay of the fluorescence light intensity with time is used to determine the time evolution of the distribution of intramolecular distances, has been utilized to study the folding of the small protein, monellin. It is seen that when folding begins, about one-third of the protein molecules collapse into a molten globule state (I_{MG}), from which they relax by continuous further contraction to transit to the native state. The larger fraction gets trapped into a metastable misfolded state. Exit from this metastable state occurs *via* collapse to the lower free energy I_{MG} state. This exit is slow, on a time scale of seconds, because of activation energy barriers. The trapped misfolded molecules as well as the I_{MG} molecules contract continuously and slowly as structure develops. A phenomenological model of Markovian evolution of the polymer chain undergoing folding, incorporating these features, has been developed, which fits well the experimentally observed time evolution of distance distributions. The observation that the “wrong turn” to the misfolded state has not been eliminated by evolution belies the common belief that the folding of functional protein sequences is very different from that of a random heteropolymer, and the former have been selected by evolution to fold quickly.

© 2019 Elsevier Ltd. All rights reserved.

Introduction

Polypeptide chains undergo a collapse in size, as they fold into functional, structured proteins [1,2]. Polypeptide-chain collapse need not always involve specific structure formation: many intrinsically disordered proteins (IDPs) remain devoid of specific structure even in their collapsed forms [3–5]. Indeed, for many proteins, it has now been shown that polypeptide chain collapse precedes specific structure formation [6–10]. Chain contraction then continues during the main folding reactions. Understanding the dynamics of the transition of the polypeptide chain from an expanded to a compact form is useful in multiple contexts such as (a) for understanding how proteins begin to fold [1,2,11], (b) for understanding how IDPs may form phase-

separated functional liquid droplets in the cell [12–14], and (c) for understanding the aggregation propensities of proteins [15–18].

Understanding the nature of the polypeptide chain collapse reaction that initiates the formation of the specific native structure during folding is important, because the degree to which structured conformations are populated during the process will dictate how subsequent folding will proceed [2,19–22]. The unfolded (U) state ensemble is structurally heterogeneous [23–26]: molecules may differ substantially in the extent of residual structure they possess, in the strength of electrostatic interactions present, and in the isomerization of X-Pro bonds. An important question, therefore, is whether all molecules collapse initially or whether only a sub-population of unfolded molecules undergoes a significant

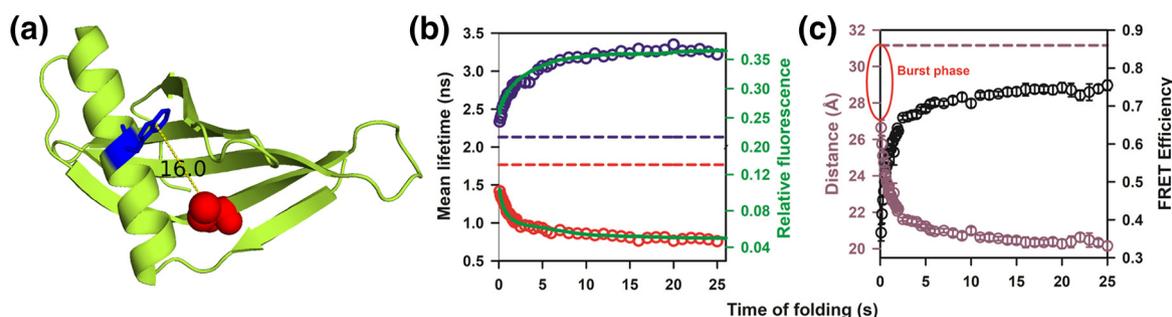


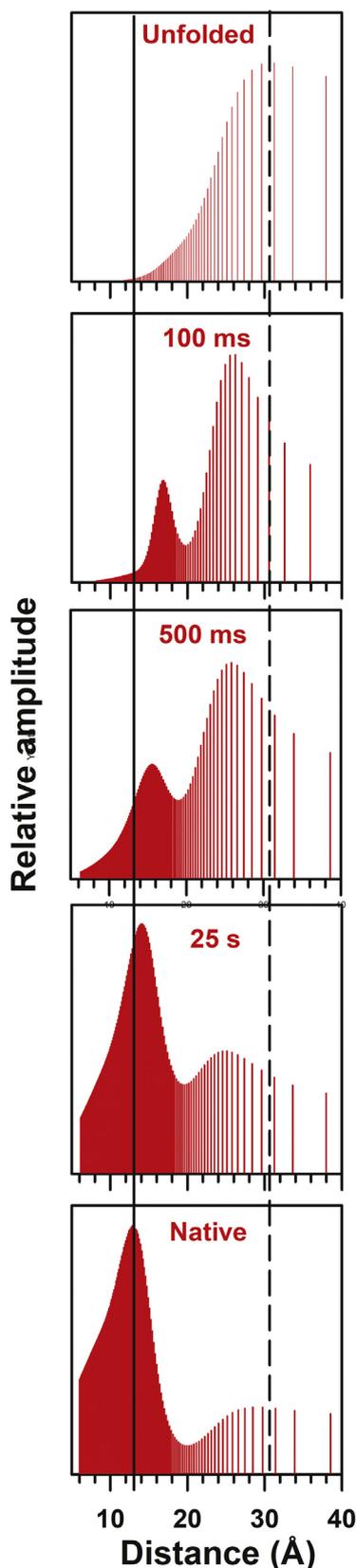
Fig. 1. Mapping collapse in the core of the protein during folding using FRET as a probe. (A) The location of the residues used as donor (Trp19, blue sticks) and acceptor (TNB attached to Cys42, red spheres) is shown in the structure of single-chain monellin (PDB entry 1IV7), drawn using the program PyMOL. (B) Kinetics of folding of unlabeled and TNB-labeled W19C42 MNEI in 0.4 M GdnHCl at pH 8 and 25 °C, monitored by steady-state and time-resolved FRET. The change in the mean fluorescence lifetime of Trp19 during refolding is shown for the unlabeled (blue open circles) and TNB-labeled (red open circles) variants (left y axis.) The dashed horizontal lines correspond to the mean lifetime for the unfolded state in 4 M GdnHCl, for the unlabeled (blue) and TNB-labeled (red) variants. The green solid lines show the changes in the fluorescence intensity of Trp19 (in the absence and presence of acceptor, TNB) during folding (right y axis). (C) Dependence of the FRET efficiency (right y axis) and corresponding distance (left y axis) between Trp19 and TNB (attached to Cys42) on the folding time. The dashed horizontal line corresponds to the average distance for the unfolded state in 4 M GdnHCl. The initial unstructured collapse of the protein chain, called the burst phase, is over within 100 ms of the initiation of folding. Error bars represent the standard errors of measurements from three independent double kinetics experiments.

contraction in size. If all molecules do not rapidly collapse, is it because of a single large free energy barrier [27–31], or is it because of many small distributed barriers on a rugged free energy landscape [32–35]? While several equilibrium studies have suggested that conformational changes during folding and unfolding occur incrementally [36–38], very few kinetic studies have been able to directly determine the nature of the barriers that slow down these reactions [39,40], or of the specificity of the collapse reaction [23,33,41–44].

Like the expanded U state, the initially collapsed form is also likely to be structurally heterogeneous, with both native and non-native, non-covalent bonds (hereafter referred to as interactions) present in some but not all molecules [45–47]. This heterogeneity would manifest itself in different sub-populations of molecules differing in both native and non-native interactions in distinct segments of their structures. Non-native interactions have now been detected in the initially collapsed forms of several proteins [47–50], as they have been in later folding intermediates [21]. In some cases, they have been shown to slow down subsequent folding [29,51–53]. Theoretical studies have provided a better understanding of the role of non-native interactions in protein folding [54–56]. Importantly, it has been suggested that the presence or absence of non-native contacts might not have an influence on the mechanism or sequence of native structure formation during folding, but would more likely affect the kinetics of folding by affecting the partitioning of folding trajectories into paths containing misfolded/non-native kinetic traps [54]. Understanding the

origin and nature of structural heterogeneity present in initially collapsed forms as well as the evolution of this heterogeneity during folding is important, because structurally distinct sub-populations of molecules would be expected to also take different pathways to the folded state [57]. Unfortunately, while measurements of the time scale of initial chain collapse have been possible for some time [58,59], they have been able to provide only glimpses of the conformational heterogeneity present in the initial collapsed ensemble [46–49].

There is some uncertainty about the extent of initial chain collapse during folding because of an apparent discrepancy in the results obtained using two different probes, small-angle x-ray scattering (SAXS) and fluorescence resonance energy transfer (FRET), to monitor collapse [60–64]. One issue is that the parameter determined directly from SAXS measurements is the radius of gyration (R_g), while the parameter directly determined from FRET measurements, using both ensemble [22,33,42,46,49,65,66] and single-molecule (smFRET) [25,26,34,45,67–69] methods, is a donor–acceptor distance (R_{da}). While these will generally both increase or both decrease with a change in solvent conditions, the fractional changes need not be equal. A direct comparison between these two measures of molecular size requires assumptions to be made about the shape and geometry of the polymer chain [63,70], which may not be correct. Second, hydrophobic dye adducts used in smFRET measurements may affect the collapse and the behavior may not be same as the unlabeled protein in some [43], but not all [42] cases.



Third, structural heterogeneity is present in the unfolded state ensemble in folding conditions [10,29–31,34–37,48–53], which arises because different fractions of the population of molecules have collapsed to various extents in distinct segments. This could lead to a given R_{da} value being consistent with several R_g values [70–73] and could lead to different time dependences of R_g and R_{da} . Clearly understanding the nature and time evolution of this structural heterogeneity is important from several different perspectives.

The small protein monellin (MNEI) (Fig. 1A) has proven to be of great utility for understanding basic aspects of folding and unfolding reactions, in both experimental [10,20,38,39,43,49,74–79] and computational [80,81] studies. In particular, it has been shown that its folding begins with a non-specific chain collapse in which the average diameter has decreased by ~35%, and some non-native interactions have formed [49]. Subsequent stepwise consolidation of the collapsed globule occurs without a concomitant decrease in average diameter and leads to the formation of a kinetic molten globule intermediate (I_{MG}) at 1 ms of folding. I_{MG} is devoid of the helix present in native MNEI [10]. The folding of I_{MG} to native protein occurs slowly along multiple pathways [76,78] and is accompanied by a decrease in average size [10] of the molecules. Hydrogen exchange–mass spectrometry (HX-MS) studies have suggested that secondary structure changes can happen gradually during the major folding reactions, depending on the folding conditions [75]. Little is, however, known about the heterogeneity in size of the initial collapsed forms (whether the distribution of sizes in the population is unimodal or is limited by a barrier, because the ensemble-averaging spectroscopic methods used so far to probe the earliest folding events lacked the capability to answer these questions.

One technique that can distinguish and also quantify the sub-populations of different conformations present together is time-resolved fluorescence resonance energy transfer (trFRET) measurements [82,83], especially when analyzed by the maximum entropy method (MEM)-based analysis [36,38,39].

Fig. 2. Evolution of distance distributions as a function of folding time (A). Experimentally derived distance distributions for representative time points of the folding reaction. The top-most panel corresponds to the unfolded (U) state, the bottom-most panel corresponds to the distance distribution of the native (N) state, and the middle panels correspond to intermediate time points, as described on each panel, during folding following a 4 to 0.4 M GdnHCl jump. The vertical solid and dashed black lines indicate the peak positions of the fluorescence lifetime distributions corresponding to the N state and U state, respectively.

Fluorescence intensity decays of a FRET donor are measured using a double-kinetics setup [39,84], at different times of folding, both in the absence and presence of a FRET acceptor. The observed fluorescence decay profiles are fitted to a sum of exponentials using a MEM algorithm. The distribution of decay times is converted into a distribution of distances between the donor and acceptor fluorophore using the Forster relation [36]. Since fluorescence intensity decays occur in the nanosecond time domain, much faster than the time scale associated with conformational fluctuations, the use of trFRET enables the changes in the distribution of distances separating donor and acceptor, in the population of molecules, to be monitored as a function of the time of folding.

Materials and Methods

In this study, a custom-built double kinetics setup (Fig. S1) was used to monitor simultaneously the changes in the fluorescence lifetime as well as in the steady-state fluorescence intensity of a tryptophan residue (W19) that acts as the FRET donor, in the absence and presence of a FRET acceptor, thionitrobenzoate (TNB), attached to C42, which accompany the folding of MNEI (Fig. 1A). The W19–C42TNB FRET pair maps the core of the protein (Fig. 1A). The protein was first unfolded in 4 M guanidine hydrochloride (GdnHCl). Folding from the unfolded (U) state to the native (N) state was initiated by reducing the GdnHCl concentration to 0.4 M using a stopped-flow mixer. The trFRET measurements determined the distribution of distances separating W19 from C42-TNB, as a function of the time of folding (Fig. 2). Fluorescence lifetime distributions of both the unlabeled and the TNB labeled variants were determined for consecutive acquisition time windows of 100 ms by fitting the

observed fluorescence decay profiles to sums of exponentials using the MEM algorithm [36,38,39]. The lifetime distributions of the TNB-labeled variant at different times of folding were then converted to FRET efficiency distributions by using the peak lifetime values obtained for the corresponding unlabeled variant (see Eq. 6 in Materials and Methods, SI). The FRET efficiency distributions were then transformed into distance distributions using the well-known Forster relation (see Eq. 7 in Materials and Methods, SI). A detailed description of all the reagents and methods used in this study, as well as of the data analysis is given in the Materials and Methods section of the Supplementary Information file.

Results and Discussion

Observation of heterogeneity in the product of the initial collapse reaction

In an earlier equilibrium unfolding study, it was established that the secondary structures and stabilities of the unlabeled and TNB-labeled mutant proteins used in the current study are similar [38]. In the current study, it was found that the kinetics of folding of the unlabeled and TNB-labeled proteins are also similar when monitored by measurement of far-UV circular dichroism (CD) and steady-state fluorescence intensity (Fig. S2). It is important to note that the TNB adduct introduced as the FRET acceptor is small and charged, and it was found to have only a weak effect on the structure and folding dynamics of the labeled protein. In earlier ensemble FRET-monitored studies of the folding of monellin [10,38,77], the introduction of the Trp–TNB FRET pair in different segments of the protein structure had

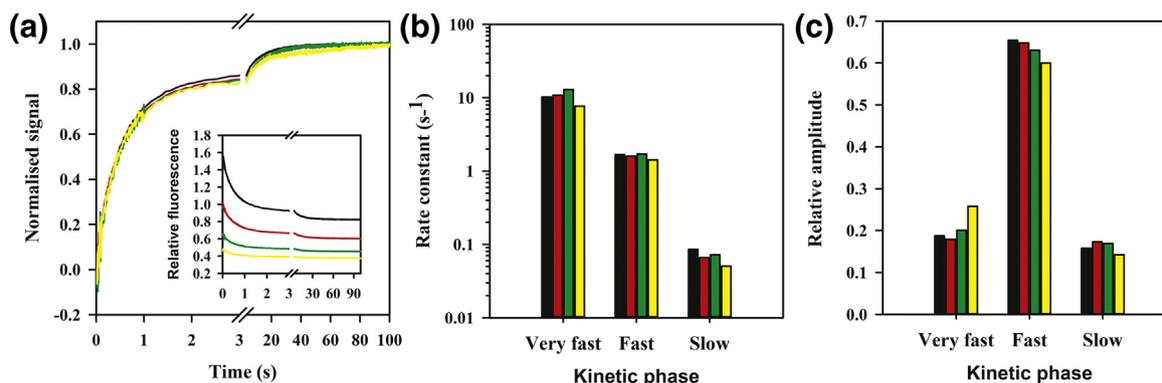


Fig. 3. The folding kinetics are independent of protein concentration. Panel A shows the kinetic traces of the folding of the TNB-labeled protein, obtained for a jump in GdnHCl concentration from 4 to 0.4 M. The protein concentrations used were 2.25 (yellow), 4.5 (green), 9 (red), and 18 (black) μM . Panels B and C show the values of the rate constants of folding and their relative amplitudes, respectively, which are obtained when the traces are fitted to a sum of three exponentials and which are characterized as very fast, fast, and slow phases.

little, if any, effect on the structure, stability, and folding kinetics. Similar results had also been obtained previously when the Trp–TNB FRET pair had been used to carry out multi-site, ensemble FRET studies of the folding of other proteins [33,35,36,47,66,85,86]. In general, it appears that small fluorophore adducts perturb protein conformational dynamics to only a small extent [2,70,87], in contrast to the large Alexa dyes typically used for smFRET measurements [62,88].

Figure 1B shows that the time courses of the fractional increases in the mean lifetime and fluorescence intensity, both simultaneously monitored, were identical. This was true for both the unlabeled and labeled proteins. This validates the accuracy of the data collection in the double-kinetic experiments (see Materials and Methods, SI for details). When the data in Fig. 1B was transformed into FRET efficiency, and then into average distance separating the FRET pair, an initial sharp (burst phase) decrease in mean size was observed: the fractional change in mean size at 100 ms of folding was $39\% \pm 6\%$ of the overall change in size from the U to the N state (Figs. 1C, 3C, and S3). It was found that when the fit through the plot of distance *versus* time of folding was constrained to start at the distance corresponding to that in the U state, the resultant fit was poor compared to the free fit (Fig. S4B), supporting the identification of the burst phase decrease in mean size as a separate kinetic phase corresponding to an initial chain collapse process.

In previous steady-state FRET studies [10,49] of the initial collapse reaction of MNEI, the use of multiple FRET pairs distributed throughout the molecule had shown that the average contraction of different intramolecular distances at 1 ms of folding was $35\% \pm 9\%$, with individual segments contracting to between 20% and 47%. Hence, the $39\% \pm 6\%$ decrease observed in the mean distance separating the W19–C42TNB FRET pair is unlikely to be a decrease for only this pair but would represent an overall reduction in size in the other measures of the size of the protein, including its radius of gyration. It should be noted that the $\sim 40\%$ reduction in mean size may be achieved by all molecules undergoing a comparable reduction in size, or by a smaller fraction (40%) becoming much smaller, and the rest (60%) remaining roughly unchanged. The latter scenario is referred to as heterogeneous collapse. This issue could be investigated experimentally by trFRET measurements. The distance distributions (Fig. 2) obtained from MEM analysis (see Materials and Methods, SI for details) of the double kinetics data revealed that at the first observable time point (100 ms) after the initiation of the folding reaction, the distribution of intramolecular distances had become bimodal. The major component ($68\% \pm 8\%$) of the bimodal distribution corresponded to the distribution of distances

in U-like molecules, named as U_x sub-ensemble hereafter, that had contracted by only about 14% compared to the size of U (31.6 Å). The minor component ($32\% \pm 8\%$) of the bimodal distribution corresponded to the distribution of distances in N-like molecules, whose size was about 48% smaller than U, but still 23% larger than that of N (13.8 Å). The N-like molecules had clearly undergone a significant chain collapse at 100 ms of folding.

The ensemble-averaged size had been shown to reduce by $35\% \pm 9\%$ within 37 μ s of the initiation of the refolding reaction, which does not change further during the first millisecond and only by an additional $\sim 5\%–10\%$ at 100 ms of folding [10,49]. It was found in this study (Figs. 2, 3, and S7) that $32\% \pm 8\%$ of the molecules had collapsed to a mean size that was only about 23% larger than that of N at 100 ms of folding in 0.4 M GdnHCl, and that the remaining $68\% \pm 8\%$ of the molecules had contracted in mean size only by 14% at this time. These observations agree very well with the previous measurements of the reduction in ensemble-averaged size. These results indicated that very little, if any, additional decrease in chain dimensions occurred in the 100 ms following the initial collapse reaction that was essentially complete within about 40 μ s of the start of the reaction.

The collapsed intermediate has molten globule-like properties

A previous study [10] of folding, under very similar but not identical conditions, had provided no evidence for secondary structure formation at 1 ms of folding and had shown that $<5\%$ of the native-like secondary structure is formed at 100 ms of folding. In the current study, the folding of the labeled protein was also monitored by measurement of CD using stopped-flow mixing (Fig. S5) under conditions identical to those used for the steady-state fluorescence intensity and trFRET measurements. It was observed that only about 2% of the total change in CD that took place during the U to N reaction had occurred at 100 ms of folding in 0.4 M GdnHCl. If this change in CD occurred only during the collapse of $32\% \pm 8\%$ of the molecules to N-like forms and not during the contraction of U to U_x , it would mean that the collapsed N-like forms possessed only about 5% of the N-state secondary structure. Clearly, very little significant structure formed at 100 ms of folding. Finally, from the amplitude-weighted effective rate constant of folding monitored by the measurement of either steady-state fluorescence intensity or CD (see Materials and Methods, SI and Fig. S10B), it was estimated that only about 3% of the total extent of structure formation had occurred at 100 ms of folding. Hence, the $\sim 40\%$ reduction in the ensemble-averaged size seen at 100 ms of folding in both the current and previous [10,49] studies was

unlikely to have been driven by the structure-forming (folding) reactions that followed initial chain collapse. Importantly, the collapsed forms at 100 ms appeared to be only marginally more structured (if at all) than the collapsed kinetic molten globule form, I_{MG} , that is populated at 1 ms, and which is devoid of native-like secondary structure [10] (Fig. S5). Hence, the ensemble of molecules with a collapsed N-like structure may be identified with the I_{MG} sub-ensemble defined earlier.

Collapse to the molten globule intermediate is an activated process

It was found that at 100 ms of folding, $68\% \pm 8\%$ of the molecules belong to U_X sub-ensemble, whose size was only slightly smaller than that of U (Fig. 2). As the folding reaction progressed, there was a reduction in the population of the U_X sub-ensemble and a corresponding increase in the population of the N-like molecules that constitute the I_{MG} sub-ensemble. (Figs. 2 and 4A). This observation suggested that the U_X sub-ensemble of molecules relaxed (folded) to the I_{MG} sub-ensemble, through a slow activated process that resulted in a major reduction in the size of the molecules. Since U_X encounters a substantial barrier in order to form the native state, it can be called as a misfolded state.

It was important to verify that both U_X and I_{MG} , populated early during folding, are monomeric in nature. For this, the kinetics of folding was monitored by the measurement of steady-state fluorescence over a range of protein concentrations (~ 2 to $18 \mu\text{M}$). If the U_X or I_{MG} ensembles were to consist of dimeric or multimeric protein molecules, then the folding kinetics

would have been different at various protein concentrations. The observation that the apparent rate constants and relative amplitudes of each of the three phases of folding are independent of protein concentration, for the TNB-labeled (Fig. 3) and unlabeled protein variants (Fig. S8), strongly suggests that transient oligomerization of the protein molecules is not an important factor in the kinetics of folding.

The kinetics of evolution of the U_X sub-ensemble to the I_{MG} sub-ensemble in a barrier-limited process was found to be describable by a double exponential equation (Fig. 4A, inset, and Fig. S9), suggesting that the two sub-ensembles were separated by two free energy barriers. At present, the structural origin of the free energy barriers that separate the U_X and I_{MG} sub-ensembles cannot be established. It could be that the molecules in the U_X sub-ensemble possess a few non-native interactions that must be broken before they can evolve to I_{MG} , or that a few specific native interactions must form before the transformation can occur. It should be noted that these non-native interactions are unlikely to have been present in the U state from which the folding reaction was initiated, and hence, the I_{MG} sub-ensemble could be initially populated to the extent of $32\% \pm 8\%$ within 100 ms of folding.

The presence of two free energy barriers separating the U_X and I_{MG} sub-ensembles suggests that the U_X sub-ensemble is heterogeneous, and consist of two sub-populations of molecules that transform into each other not faster than they transform to I_{MG} , so that each subpopulation of U_X evolves independently to I_{MG} . One possibility is that the two subpopulations of molecules differ in conformation at the two peptidyl-prolyl bonds that are known to be present in the *cis* form in the N state, and which

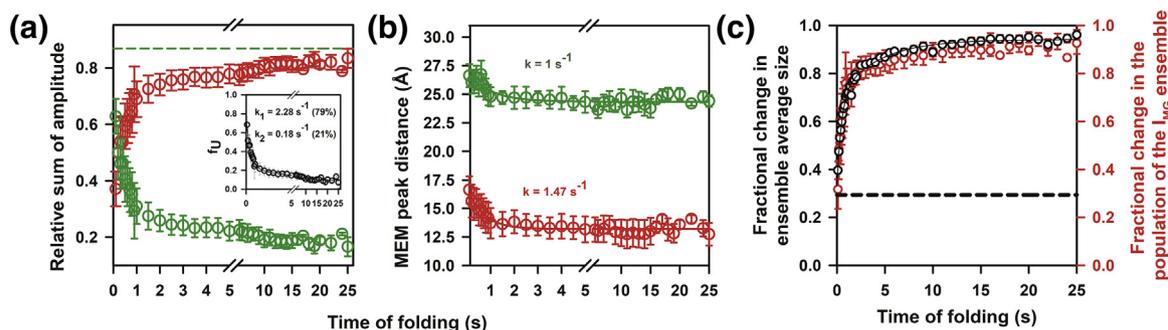


Fig. 4. Co-occurrence of barrier-limited collapse and barrier-less contraction. (A) Barrier-limited collapse monitored by the relative sum of amplitudes under the N-like (I_{MG} , red symbols) and U-like (U_X , green symbols) distributions for the TNB-labeled protein. The relative sum of amplitudes for each population is the sum of amplitudes of the distribution for that population divided by the sum of amplitudes of the distributions for both the N-like and U-like populations. The unfolded protein baseline is shown as a dashed line. The inset in panel A corresponds to the fractional population of the U_X sub-ensemble relative to the N ($f_U = 0$) and U ($f_U = 1$) states. (B) Barrier-less contraction of the U_X (green symbols) and I_{MG} (red symbols) sub-ensembles, monitored by the distance calculated from peak FRET efficiency of the U-like and N-like distributions, respectively, for all time points of the folding reaction. (C) Dependence on the folding time, of the ensemble-averaged size (distance) obtained from mean lifetime of double kinetics data (left y axis), and of the fractional change in the population of the N-like (I_{MG}) sub-ensemble (right y axis). The dashed horizontal line corresponds to the ensemble-averaged distance measured at 1 ms of folding reaction using steady-state FRET measurements. The error bars shown are the standard errors obtained from three independent double kinetics experiments.

would be expected to be predominantly in the *trans* form in the U state and U_X sub-ensembles [76].

The observation that the evolution of the U_X sub-ensemble to the I_{MG} sub-ensemble is an activated process suggests strongly that the initial collapse of the U ensemble, which was observed to be essentially complete at 37 μ s, in an earlier steady-state FRET study of folding [49], is also an activated process. Also, a quantitative correspondence between the folding time dependence of fractional change in the ensemble-averaged size and the population of the I_{MG} sub-ensemble (Fig. 4C) suggests that the accumulation of I_{MG} is the major contributor to the overall reduction in dimensions throughout the folding reaction including the initial collapse at 1 ms. An earlier trFRET study had shown, though only qualitatively, that initial chain collapse during the folding of cytochrome *c* is also barrier-limited [40].

Observation of glassy behavior in the contraction of both the sub-populations

It was found that the molecules in both the U_X and I_{MG} sub-ensembles undergo slow continuous contraction in size, which is concomitant with U_X evolving to I_{MG} (Fig. 2). The contraction in size was apparent in the movements of the positions of the peaks of the distributions corresponding to both sub-ensembles (Fig. 4B). The I_{MG} sub-ensemble evolved gradually to form the more compact N ensemble, while the U_X sub-ensemble evolved gradually to form the more compact U_X^* . In many respects, the evolution of I_{MG} to the more compact N, and of U_X to the more compact U_X^* , is similar to homopolymer collapse, which has been predicted to be continuous when the homopolymer is transferred from a good to a bad or theta solvent [1,89–91]. The gradual contraction of molecules within both the U_X and I_{MG} sub-ensembles was, however, much slower than expected for the coil to globule transition of a homopolymer. An order of magnitude calculation of the effective viscosity for the observed gradual contraction in both the sub-ensembles gives a value of the effective viscosity that is 10^9 times the viscosity of water at room temperature (see Materials and Methods, SI). This large effective viscosity is a measure of the slow glassy dynamics of folding in this system. The origin of this large viscosity is not well understood [89,92–94]. It is likely that because of the ruggedness of the free energy landscape of the protein and because the rate of formation of the first few native contacts is small because of entropic barriers, contraction is slowed down. These specific interactions remain to be identified.

Role of denaturant in the collapse and folding of MNEI

The extent to which the I_{MG} sub-ensemble was populated at 100 ms of folding was found to increase

non-linearly with a decrease in the GdnHCl concentration in which folding was initiated (Fig. S10B). The fractional weight of the I_{MG} sub-ensemble decreased from 0.68 to 0.26, as the concentration of GdnHCl was increased from 0.1 to 0.6 M GdnHCl. This suggests that the initial collapse to the I_{MG} sub-ensemble plays a major role in mediating the effect of denaturant on the folding of the protein. The size of the molecules in both the U_X and I_{MG} sub-ensembles was found to also decrease with a reduction in GdnHCl concentration (Fig. S10C). This dependence appeared to be linear in nature and was similar to what is expected for any homopolymer when the solvent quality is improved [91,95]. A similar dependence had been observed earlier for the contraction of the unfolded states of globular proteins, as well as of IDPs, upon a reduction in denaturant concentration, in various unfolding studies [26,38,69,96,97]. The contraction seen in the I_{MG} sub-ensemble under more stabilizing folding conditions had, however, not been observed in earlier kinetic studies, possibly because of the limited sensitivity of the fluorophores used, in this short distance regime.

A phenomenological coarse-grained Markov evolution model can explain the experimental observations in a quantitative manner

The characteristic time (~ 1 s) of continuous contraction of molecular size in both the U_X and I_{MG} sub-ensembles, as well as of the evolution of the U_X sub-ensemble to the I_{MG} sub-ensemble, corresponded well with the apparent rate constant of the major kinetic phase of the folding reaction [76,78]. This observation suggested that the continuous contraction is accompanied by or induces specific native structure formation. This suggests that after the initial collapse that leads to the formation of the structure-less I_{MG} sub-ensemble over the first millisecond of folding [49], native structure formation occurs in a gradual manner, as the I_{MG} sub-ensemble undergoes continuous chain compaction to evolve to N. A previous study of the unfolding of MNEI had shown that unfolding occurred by gradual swelling of the molecules along two pathways, that the swelling was accompanied by an activated process through which molecules were channeled to the productive unfolding pathway, and that loss of structure accompanied both types of processes [39]. Slow swelling during unfolding could be described by a model based on Rouse-like polymer chain dynamics, but with a very small effective monomer diffusion constant. Also, the swelling was taken to be effectively unidirectional, parameterized by a Gaussian whose mean and width were chosen to best fit the experimental data. A variation of this model (see Materials and Methods, SI) adequately describes the data in Fig. 2: distance distributions obtained assuming Rouse-like polymer chain dynamics

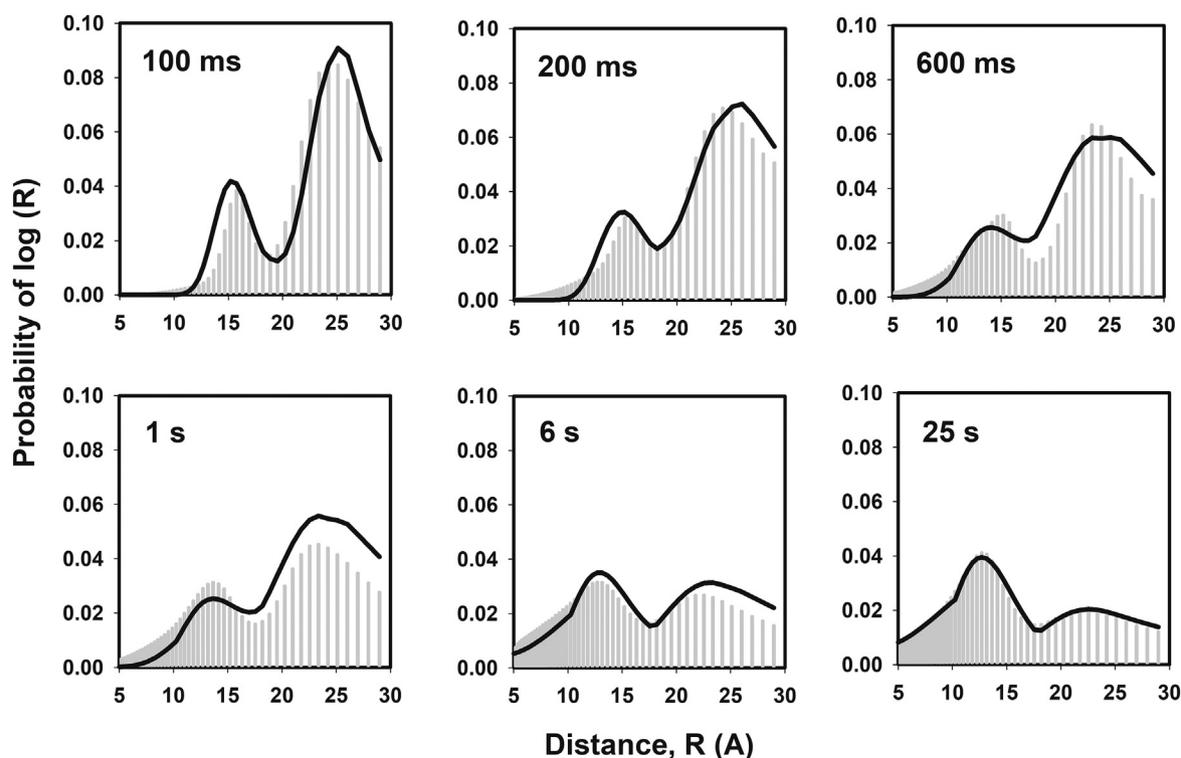


Fig. 5. Probability distributions of distances obtained using a coarse-grained Markov evolution model. A comparison of experimental data (gray vertical bars) and simulated data (black solid line) obtained using a coarse-grained Markov evolution model (see Materials and Methods, SI for details). Different panels correspond to different times of the folding reaction as described on the top left of each panel.

during folding could satisfactorily describe the experimentally observed distance distributions (Fig. S11B; Table S1). Hence, the folding process is described as a series of diffusive structural changes leading to the formation of the N state. HX-MS studies on MNEI had also earlier shown that secondary structure is formed (and lost) in a gradual manner [75].

In this study, the Markovian evolution model has been reformulated to generalize it and to extend its applicability beyond the Rouse-like model. The phase space of conformations of molecules is discretized into a finite number of disjoint classes (about 100). Each class is characterized by a small number of macroscopic variables like the mean radius of gyration and distance from the native conformation. Within a class, the relative weights of different conformations are assumed to be the Boltzmann weights in the equilibrium ensemble. It is assumed that the time evolution of the system is Markovian, and the transition probabilities satisfy the detailed balance condition. Then, the progress of the reaction is described by how the probabilities of different classes evolve in time, as the reaction proceeds. The classes have been characterized using two parameters: the extent of contraction and the degree of nativeness (Fig. 6). Upon initiation of

the folding reaction, U contracts and collapses to U_X and I_{MG} . As folding progresses, U_X and I_{MG} undergo gradual contraction to form U_X^* and N, respectively. Simultaneously, the U-like sub-ensemble converts to the N-like sub-ensemble by an activated process. It is to be noted that the transitions from U to U_X , U_X to U_X^* , and I_{MG} to N are downhill or gradual in nature. However, a significant free energy barrier is encountered in the formation of I_{MG} , either from U or from U_X . The landscape (Fig. 6) is a qualitative representation consistent with the experimental data, and with both the models (Rouse-like polymer chain model and coarse-grained Markov evolution model) used for describing the kinetic data for folding.

The values of the transition probabilities are determined by fitting the data from trFRET observations. The detailed description of the modeling procedure is given in the supplementary text (Materials and Methods, SI). An excellent quantitative agreement between the experimentally derived distributions and the distributions obtained from the model is seen, at different times of folding (Fig. 5), and this picture in terms of evolution in the phase space helps in visualizing the population evolution during folding (Fig. 6).

Interestingly, folding behavior very similar to that seen in the current study was found in an early

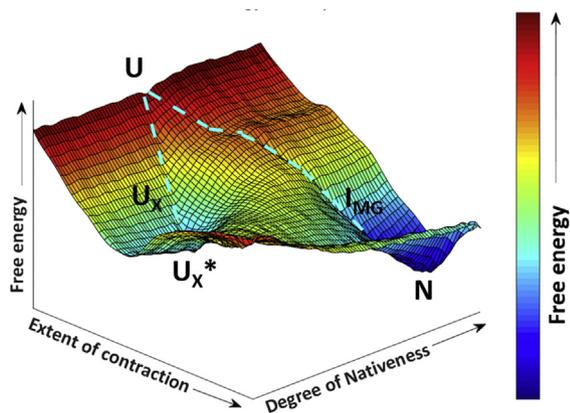


Fig. 6. Schematic free energy landscape for describing collapse and folding of single-chain monellin. The free energy of folding is plotted as a function of two structural parameters, the extent of contraction, and the degree of nativeness. All the three axes have arbitrary units. Unfolded state is shown as “U,” “ U_X to U_X^* ” constitute the U state sub-ensemble, and “ I_{MG} to N” constitute the N state sub-ensemble. The major pathways are denoted by turquoise dashed lines.

Monte Carlo study of folding. In that study, a lattice model of a rather short-chain heteropolymer was studied on a simple two-dimensional lattice [98]. That study found a compact intermediate state with many non-native hydrophobic interactions. The same study also found multiple folding pathways, which have been observed experimentally for the folding of MNEI [76,78].

Resolving heterogeneity in polypeptide chain collapse

The trFRET technique is ideally suited to probe the nature of the initial events, especially chain collapse, during protein folding, as it allows populations of different co-existing conformations of a protein to be identified and quantified [36,82]. While smFRET studies can in principle provide similar information, their application has been limited largely to equilibrium studies [5,68,69,96], because kinetic smFRET experiments [67,99,100] are technically challenging. The HX-MS methodology is a complimentary method to distinguish between different coexisting conformations, but it does not provide direct information on chain collapse [101].

Much of the debate [1,29,30,35] about the nature of polypeptide chain collapse during protein folding has arisen because of confusion between the interpretations of the results of measurements that correspond to ensemble averages of different physical quantities. For example, if the average size X is determined as the mean squared size, its value will differ from the value deduced from measuring the average value of $\exp(-q^2X^2)$, which is the quantity measured by SAXS. While

steady-state FRET has shown that folding commences with non-specific hydrophobic collapse [6,33], that there are both specific and non-specific components to collapse [23,42], and that different segments of the protein collapse in a non-synchronized manner [42], it could not provide information on whether the observed changes happened in all molecules or were restricted to a sub-population of molecules. In contrast, the current study shows that trFRET measurements can not only distinguish between expanded and collapsed sub-populations of molecules of different size but also show that each sub-population contracts continuously as folding progresses.

Conclusions

In summary, the present study shows that two distinct kinds of changes in chain size have occurred at the earliest observable time point (100 ms), during the folding of MNEI. About a third of the molecules have collapsed to an N-like sub-ensemble, I_{MG} , whose dimensions are only slightly larger than those of the N ensemble, but which is devoid of native-like secondary structure. The remaining two-thirds of the molecules have contracted to a metastable state, called U_X here, whose size is only slightly smaller than that of U. As folding progresses from 100 ms onward, both the I_{MG} and U_X sub-ensembles undergo slow continuous contraction with time. I_{MG} evolves gradually to form N, while the U_X sub-ensemble evolves gradually to form U_X^* . The U_X sub-ensemble eventually relaxes to the I_{MG} sub-ensemble, through a slow activated process on the time scale of seconds.

Acknowledgment

We thank S.K. Jha for discussion and help in modeling the data. We thank F. Ali and T. Agarwal for help in programming. We thank R. Goluguri and other members of our laboratory, as well as M.K. Mathew and S. Gosavi, for discussion. J.B.U. is a recipient of a JC Bose National Fellowship from the Government of India. This work was funded by the Tata Institute of Fundamental Research and by the Department of Biotechnology, Government of India.

Conflict of Interest Statement: The authors declare no competing financial interests.

Appendix A. Supplementary data

SI contains detailed experimental procedures and supplementary figures. Supplementary data to this

article can be found online at <https://doi.org/10.1016/j.jmb.2019.07.024>.

Received 2 April 2019;

Received in revised form 12 July 2019;

Accepted 12 July 2019

Available online 19 July 2019

Keywords:

monellin;
collapse;
time-resolved FRET;
heterogeneity;
Markovian

Abbreviations used:

MNEI, single-chain monellin; GdnHCl, guanidine hydrochloride; FRET, fluorescence resonance energy transfer; TNB, thionitrobenzoate; MG, molten globule; SAXS, small-angle x-ray scattering.

References

- [1] J.B. Udgaonkar, Polypeptide chain collapse and protein folding, *Arch. Biochem. Biophys.* 531 (2013) 24–33, <https://doi.org/10.1016/j.abb.2012.10.003>.
- [2] D. Thirumalai, H.S. Samanta, H. Maity, G. Reddy, Universal nature of collapsibility in the context of protein folding and evolution, *Trends Biochem. Sci.* 461046 (2019) <https://doi.org/10.1016/j.tibs.2019.04.003>.
- [3] A.S. Morar, A. Olteanu, G.B. Young, G.J. Pielak, Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome c, *Protein Sci.* 10 (2001) 2195–2199, <https://doi.org/10.1110/ps.24301>.
- [4] S.L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, R.V. Pappu, Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions, *Proc. Natl. Acad. Sci.* 103 (2006) 16764–16769, <https://doi.org/10.1073/pnas.0608175103>.
- [5] W. Zheng, A. Borgia, K. Buholzer, A. Grishaev, B. Schuler, R.B. Best, Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment, *J. Am. Chem. Soc.* 138 (2016) 11702–11713, <https://doi.org/10.1021/jacs.6b05443>.
- [6] V.R. Agashe, M.C.R. Shastry, J.B. Udgaonkar, Initial hydrophobic collapse in the folding of barstar, *Nature.* 377 (1995) 754–757, <https://doi.org/10.1038/377754a0>.
- [7] U. Nath, J.B. Udgaonkar, Folding of tryptophan mutants of barstar: evidence for an initial hydrophobic collapse on the folding pathway, *Biochemistry.* 36 (1997) 8602–8610, <https://doi.org/10.1021/bi970426z>.
- [8] B.R. Rami, J.B. Udgaonkar, Mechanism of formation of a productive molten globule form of barstar, *Biochemistry.* 41 (2002) 1710–1716, <https://doi.org/10.1021/bi0120300>.
- [9] C. Magg, J. Kubelka, G. Holtermann, E. Haas, F.X. Schmid, Specificity of the initial collapse in the folding of the cold shock protein, *J. Mol. Biol.* 360 (2006) 1067–1080, <https://doi.org/10.1016/j.jmb.2006.05.073>.
- [10] R.R. Goluguri, J.B. Udgaonkar, Rise of the helix from a collapsed globule during the folding of monellin, *Biochemistry.* 54 (2015) 5356–5365, <https://doi.org/10.1021/acs.biochem.5b00730>.
- [11] H.S. Chan, K.A. Dill, Origins of structure in globular proteins, *Proc. Natl. Acad. Sci.* 87 (1990) 6388–6392, <https://doi.org/10.1073/pnas.87.16.6388>.
- [12] S. Milles, E.A. Lemke, Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153, *Biophys. J.* 101 (2011) 1710–1719, <https://doi.org/10.1016/j.bpj.2011.08.025>.
- [13] J.P. Brady, P.J. Farber, A. Sekhar, Y.-H. Lin, R. Huang, A. Bah, T.J. Nott, H.S. Chan, A.J. Baldwin, J.D. Forman-Kay, L.E. Kay, Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation, *Proc. Natl. Acad. Sci.* (2017) 201706197. doi:10.1073.
- [14] M.T. Wei, S. Elbaum-Garfinkle, A.S. Holehouse, C.C.H. Chen, M. Feric, C.B. Arnold, R.D. Priestley, R.V. Pappu, C. P. Brangwynne, Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles, *Nat. Chem* 9 (2017) <https://doi.org/10.1038/NCHEM.2803>.
- [15] S.A. Waldauer, O. Bakajin, L.J. Lapidus, Extremely slow intramolecular diffusion in unfolded protein L, *Proc. Natl. Acad. Sci.* 107 (2010) 1–5, <https://doi.org/10.1073/pnas.1005415107>.
- [16] L.J. Lapidus, Understanding protein aggregation from the view of monomer dynamics, *Mol. BioSyst.* 9 (2013) 29–35, <https://doi.org/10.1039/c2mb25334h>.
- [17] F. Chiti, C.M. Dobson, Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade, *Annu. Rev. Biochem.* 86 (2017) 27–68, <https://doi.org/10.1146/annurev-biochem-061516-045115>.
- [18] K.R. Srivastava, L.J. Lapidus, Prion protein dynamics before aggregation, *Proc. Natl. Acad. Sci.* 114 (2017) 3572–3577, <https://doi.org/10.1073/pnas.1620400114>.
- [19] T.R. Sosnick, M.D. Shtilerman, L. Mayne, S.W. Englander, Ultrafast signals in protein folding and the polypeptide contracted state, *Proc. Natl. Acad. Sci. U. S. A.* 94 (1997) 8545–8550, <https://doi.org/10.1073/pnas.94.16.8545>.
- [20] T. Kimura, T. Uzawa, K. Ishimori, I. Morishima, S. Takahashi, T. Konno, S. Akiyama, T. Fujisawa, Specific collapse followed by slow hydrogen-bond formation of -sheet in the folding of single-chain monellin, *Proc. Natl. Acad. Sci.* 102 (2005) 2748–2753, <https://doi.org/10.1073/pnas.0407982102>.
- [21] D.J. Brockwell, S.E. Radford, Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* 17 (2007) 30–37, <https://doi.org/10.1016/j.sbi.2007.01.003>.
- [22] M. Arai, M. Iwakura, C.R. Matthews, O. Bilsel, Microsecond subdomain folding in dihydrofolate reductase, *J. Mol. Biol.* 410 (2011) 329–342, <https://doi.org/10.1016/j.jmb.2011.04.057>.
- [23] E. Welker, K. Maki, M.C.R. Shastry, D. Juminaga, R. Bhat, H.A. Scheraga, H. Roder, Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A., *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 17681–6. doi: <https://doi.org/10.1073/pnas.0407999101>.
- [24] T.L. Religa, J.S. Markson, U. Mayor, S.M.V. Freund, A.R. Fersht, Solution structure of a protein denatured state and folding intermediate, *Nature.* 437 (2005) 1053–1056. doi: <https://doi.org/10.1038/nature04054>.
- [25] H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, B. Schuler, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule

- spectroscopy, *Proc. Natl. Acad. Sci.* 109 (2012) 16155–16160, <https://doi.org/10.1073/pnas.1207719109>.
- [26] M. Aznauryan, L. Delgado, A. Soranno, D. Nettels, J. Huang, A.M. Labhardt, S. Grzesiek, B. Schuler, Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS, *Proc. Natl. Acad. Sci.* 113 (2016) E5389–E5398, <https://doi.org/10.1073/pnas.1607193113>.
- [27] D. Stigter, D.O.V. Alonso, K.A. Dill, Protein stability: electrostatics and compact denatured states, *Proc. Natl. Acad. Sci. U. S. A.* 88 (1990) 4176–4180.
- [28] O.B. Ptitsyn, V.N. Uversky, The molten globule is a third thermodynamical state of protein molecules, *FEBS Lett.* 341 (1994) 15–18, [https://doi.org/10.1016/0014-5793\(94\)80231-9](https://doi.org/10.1016/0014-5793(94)80231-9).
- [29] T.R. Sosnick, L. Mayne, R. Hiller, S.W. Englander, The barriers in protein folding, *Nat. Struct. Biol.* 1 (1994) 149–156, <https://doi.org/10.1038/nsb0394-149>.
- [30] M.C.R. Shastry, H. Roder, Evidence for barrier-limited protein folding kinetics on the microsecond time scale, *Nat. Struct. Biol.* 5 (1998) 385–392, <https://doi.org/10.1038/nsb0598-385>.
- [31] S.J. Hagen, W.A. Eaton, Two-state expansion and collapse of a polypeptide, *J. Mol. Biol.* 301 (2000) 1019–1027, <https://doi.org/10.1006/jmbi.2000.3969>.
- [32] M.J. Parker, S. Marqusee, The cooperativity of burst phase reactions explored, *J. Mol. Biol.* 293 (1999) 1195–1210, <https://doi.org/10.1006/jmbi.1999.3204>.
- [33] A. Dasgupta, J.B. Udgaonkar, Evidence for initial non-specific polypeptide chain collapse during the refolding of the SH3 domain of PI3 kinase, *J. Mol. Biol.* 403 (2010) 430–445, <https://doi.org/10.1016/j.jmb.2010.08.046>.
- [34] E. Sherman, G. Haran, Coil-globule transition in the denatured state of a small protein, *Proc. Natl. Acad. Sci.* 103 (2006) 11539–11543, <https://doi.org/10.1073/pnas.0601395103>.
- [35] K.K. Sinha, J.B. Udgaonkar, Barrierless evolution of structure during the submillisecond refolding reaction of a small protein, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 7998–8003, <https://doi.org/10.1073/pnas.0803193105>.
- [36] G.S. Lakshminanth, K. Sridevi, G. Krishnamoorthy, J.B. Udgaonkar, Structure is lost incrementally during the unfolding of barstar, *Nat. Struct. Biol.* 8 (2001) 799–804, <https://doi.org/10.1038/nsb0901-799>.
- [37] A. Narayan, L.A. Campos, S. Bhatia, D. Fushman, A.N. Naganathan, Graded structural polymorphism in a bacterial thermosensor protein, *J. Am. Chem. Soc.* 139 (2017) 792–802, <https://doi.org/10.1021/jacs.6b10608>.
- [38] S. Bhatia, G. Krishnamoorthy, J.B. Udgaonkar, Site-specific time-resolved FRET reveals local variations in the unfolding mechanism in an apparently two-state protein unfolding transition, *Phys. Chem. Chem. Phys.* 20 (2018) 3216–3232, <https://doi.org/10.1039/C7CP06214A>.
- [39] S.K. Jha, D. Dhar, G. Krishnamoorthy, J.B. Udgaonkar, Continuous dissolution of structure during the unfolding of a small protein, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 11113–11118, <https://doi.org/10.1073/pnas.0812564106>.
- [40] S.V. Kathuria, C. Kayatekin, R. Barrea, E. Kondrashkina, R. Graceffa, L. Guo, R.P. Nobrega, S. Chakravarthy, C.R. Matthews, T.C. Irving, O. Bilsel, Microsecond barrier-limited chain collapse observed by time-resolved FRET and SAXS, *J. Mol. Biol.* 426 (2014) 1980–1994. doi:<https://doi.org/10.1016/j.jmb.2014.02.020>.
- [41] M. Arai, E. Kondrashkina, C. Kayatekin, C.R. Matthews, M. Iwakura, O. Bilsel, Microsecond hydrophobic collapse in the folding of *Escherichia coli* dihydrofolate reductase, an α -type protein, *J. Mol. Biol.* 368 (2007) 219–229, <https://doi.org/10.1016/j.jmb.2007.01.085>.
- [42] K.K. Sinha, J.B. Udgaonkar, Dissecting the non-specific and specific components of the initial folding reaction of barstar by multi-site FRET measurements, *J. Mol. Biol.* 370 (2007) 385–405, <https://doi.org/10.1016/j.jmb.2007.04.061>.
- [43] T. Kimura, S. Akiyama, T. Uzawa, K. Ishimori, I. Morishima, T. Fujisawa, S. Takahashi, Specifically collapsed intermediate in the early stage of the folding of ribonuclease a, *J. Mol. Biol.* 350 (2005) 349–362, <https://doi.org/10.1016/j.jmb.2005.04.074>.
- [44] H. Fazelinia, M. Xu, H. Cheng, H. Roder, Ultrafast hydrogen exchange reveals specific structural events during the initial stages of folding of cytochrome c, *J. Am. Chem. Soc.* 136 (2014) 733–740, <https://doi.org/10.1021/ja410437d>.
- [45] B. Schuler, E.A. Lipman, W.A. Eaton, Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy, *Nature*. 419 (2002) 743–747, <https://doi.org/10.1038/nature01060>.
- [46] V. Ratner, D. Amir, E. Kahana, E. Haas, Fast collapse but slow formation of secondary structure elements in the refolding transition of *E. coli* adenylate kinase, *J. Mol. Biol.* 352 (2005) 683–699, <https://doi.org/10.1016/j.jmb.2005.06.074>.
- [47] Y. Wu, E. Kondrashkina, C. Kayatekin, C.R. Matthews, O. Bilsel, Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein, *Proc. Natl. Acad. Sci.* 105 (2008) 13367–13372, <https://doi.org/10.1073/pnas.0802788105>.
- [48] L.E. Rosen, S. V. Kathuria, C.R. Matthews, O. Bilsel, S. Marqusee, Non-native structure appears in microseconds during the folding of *E. coli* RNase H, *J. Mol. Biol.* 427 (2015) 443–453. doi:<https://doi.org/10.1016/j.jmb.2014.10.003>.
- [49] R.R. Goluguri, J.B. Udgaonkar, Microsecond rearrangements of hydrophobic clusters in an initially collapsed globule prime structure formation during the folding of a small protein, *J. Mol. Biol.* 428 (2016) 3102–3117, <https://doi.org/10.1016/j.jmb.2016.06.015>.
- [50] K. Sakurai, M. Yagi, T. Konuma, S. Takahashi, C. Nishimura, Y. Goto, Non-native α -helices in the initial folding intermediate facilitate the ordered assembly of the β -barrel in β -lactoglobulin, *Biochemistry*. 56 (2017) 4799–4807, <https://doi.org/10.1021/acs.biochem.7b00458>.
- [51] N. Schönbrunner, K.P. Koller, T. Kiefhaber, Folding of the disulfide-bonded β -sheet protein tendamistat: rapid two-state folding without hydrophobic collapse, *J. Mol. Biol.* 268 (1997) 526–538, <https://doi.org/10.1006/jmbi.1997.0960>.
- [52] H.S. Chung, S. Piana-Agostinetti, D.E. Shaw, W.A. Eaton, Structural origin of slow diffusion in protein folding, *Science*. 349 (2015) 1504–1510, <https://doi.org/10.1126/science.aab1369>.
- [53] F. Bruno Da Silva, V.G. Contessoto, V.M. De Oliveira, J. Clarke, V.B.P. Leite, Non-native cooperative interactions modulate protein folding rates, *J. Phys. Chem. B* 122 (2018) 10817–10824, <https://doi.org/10.1021/acs.jpcc.8b08990>.
- [54] T. Chen, J. Song, H.S. Chan, Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding, *Curr. Opin. Struct. Biol.* 30 (2015) 32–42, <https://doi.org/10.1016/j.sbi.2014.12.002>.
- [55] L. Huynh, C. Neale, R. Pomès, H.S. Chan, Molecular recognition and packing frustration in a helical protein, *PLoS Comput. Biol.* 13 (2017) 1–25, <https://doi.org/10.1371/journal.pcbi.1005909>.
- [56] J. Hu, T. Chen, M. Wang, H.S. Chan, Z. Zhang, A critical comparison of coarse-grained structure-based approaches

- and atomic models of protein folding, *Phys. Chem. Chem. Phys.* 19 (2017) 13629–13639, <https://doi.org/10.1039/c7cp01532a>.
- [57] J.B. Udgaonkar, Multiple routes and structural heterogeneity in protein folding, *Annu. Rev. Biophys.* 37 (2008) 489–510, <https://doi.org/10.1146/annurev.biophys.37.032807.125920>.
- [58] M. Sadqi, L.J. Lapidus, V. Munoz, How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci.* 100 (2003) 12117–12122, <https://doi.org/10.1073/pnas.2033863100>.
- [59] D. Nettels, I.V. Gopich, A. Hoffmann, B. Schuler, Ultrafast dynamics of protein collapse from single-molecule photon statistics, *Proc. Natl. Acad. Sci.* 104 (2007) 2655–2660, <https://doi.org/10.1073/pnas.0611093104>.
- [60] J. Jacob, B. Krantz, R.S. Dothager, P. Thiyagarajan, T.R. Sosnick, Early collapse is not an obligate step in protein folding, *J. Mol. Biol.* 338 (2004) 369–382, <https://doi.org/10.1016/j.jmb.2004.02.065>.
- [61] T.Y. Yoo, S.P. Meisburger, J. Hinshaw, L. Pollack, G. Haran, T.R. Sosnick, K. Plaxco, Small-angle x-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state, *J. Mol. Biol.* 418 (2012) 226–236, <https://doi.org/10.1016/j.jmb.2012.01.016>.
- [62] H.M. Watkins, A.J. Simon, T.R. Sosnick, E.A. Lipman, R.P. Hjelm, K.W. Plaxco, Random coil negative control reproduces the discrepancy between scattering and FRET measurements of denatured protein dimensions, *Proc. Natl. Acad. Sci.* 112 (2015) 6631–6636, <https://doi.org/10.1073/pnas.1418673112>.
- [63] A. Borgia, W. Zheng, K. Buholzer, M.B. Borgia, A. Schüller, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, R.B. Best, B. Schuler, Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods, *J. Am. Chem. Soc.* 138 (2016) 11714–11726, <https://doi.org/10.1021/jacs.6b05917>.
- [64] J.A. Riback, M.A. Bowman, A.M. Zmyslowski, C.R. Knoverek, J.M. Jumper, J.R. Hinshaw, E.B. Kaye, K.F. Freed, P.L. Clark, T.R. Sosnick, Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water, *Science*. 358 (2017) 238–241, <https://doi.org/10.1126/science.aan5774>.
- [65] E.V. Pletneva, H.B. Gray, J.R. Winkler, Snapshots of cytochrome *c* folding, *Proc. Natl. Acad. Sci.* 102 (2005) 18397–18402, <https://doi.org/10.1073/pnas.0509076102>.
- [66] T. Mizukami, M. Xu, H. Cheng, H. Roder, K. Maki, Nonuniform chain collapse during early stages of staphylococcal nuclease folding detected by fluorescence resonance energy transfer and ultrarapid mixing methods, *Protein Sci.* 22 (2013) 1336–1348, <https://doi.org/10.1002/pro.2320>.
- [67] E.A. Lipman, B. Schuler, O. Bakajin, W.A. Eaton, Single-molecule measurement of protein folding kinetics, *Science*. 301 (2003) 1233–1235, <https://doi.org/10.1126/science.1085399>.
- [68] E.V. Kuzmenkina, C.D. Heyes, G. Ulrich Nienhaus, Single-molecule FRET study of denaturant induced unfolding of RNase H, *J. Mol. Biol.* 357 (2006) 313–324, <https://doi.org/10.1016/j.jmb.2005.12.061>.
- [69] K. A. Merchant, R.B. Best, J.M. Louis, I.V. Gopich, W.A. Eaton, Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations., *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 1528–1533. doi:<https://doi.org/10.1073/pnas.0607097104>.
- [70] G. Fuertes, N. Banterle, K.M. Ruff, A. Chowdhury, D. Mercadante, C. Koehler, M. Kachala, G. Estrada Girona, S. Milles, A. Mishra, P.R. Onck, F. Gräter, S. Esteban-Martín, R. V. Pappu, D.I. Svergun, E.A. Lemke, Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements, *Proc. Natl. Acad. Sci.* 114 (2017) E6342–E6351. doi: <https://doi.org/10.1073/pnas.1704692114>.
- [71] H. Maity, G. Reddy, Folding of protein L with implications for collapse in the denatured state ensemble, *J. Am. Chem. Soc.* 138 (2016) 2609–2616, <https://doi.org/10.1021/jacs.5b11300>.
- [72] J. Song, G.N. Gomes, T. Shi, C.C. Gradinaru, H.S. Chan, Conformational heterogeneity and FRET data interpretation for dimensions of unfolded proteins, *Biophys. J.* 113 (2017) 1012–1024, <https://doi.org/10.1016/j.bpj.2017.07.023>.
- [73] I. Peran, A.S. Holehouse, I.S. Carrico, R.V. Pappu, O. Bilsel, D.P. Raleigh, Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions, *Proc. Natl. Acad. Sci.* 201818206 (2019) <https://doi.org/10.1073/pnas.1818206116>.
- [74] P. Malhotra, J.B. Udgaonkar, Tuning cooperativity on the free energy landscape of protein folding, *Biochemistry*. 54 (2015) 3431–3441, <https://doi.org/10.1021/acs.biochem.5b00247>.
- [75] P. Malhotra, J.B. Udgaonkar, Secondary structural change can occur diffusely and not modularly during protein folding and unfolding reactions, *J. Am. Chem. Soc.* 138 (2016) 5866–5878, <https://doi.org/10.1021/jacs.6b03356>.
- [76] A.K. Patra, J.B. Udgaonkar, Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways, *Biochemistry*. 46 (2007) 11727–11743, <https://doi.org/10.1021/bi701142a>.
- [77] S.K. Jha, J.B. Udgaonkar, Direct evidence for a dry molten globule intermediate during the unfolding of a small protein, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 12289–12294, <https://doi.org/10.1073/pnas.0905744106>.
- [78] S.K. Jha, A. Dasgupta, P. Malhotra, J.B. Udgaonkar, Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry, *Biochemistry*. 50 (2011) 3062–3074, <https://doi.org/10.1021/bi1006332>.
- [79] P. Malhotra, J.B. Udgaonkar, High-energy intermediates in protein unfolding characterized by thiol labeling under natively-like conditions, *Biochemistry*. 53 (2014) 3608–3620, <https://doi.org/10.1021/bi401493t>.
- [80] H. Maity, G. Reddy, Thermodynamics and kinetics of single-chain monellin folding with structural insights into specific collapse in the denatured state ensemble, *J. Mol. Biol.* (2017) <https://doi.org/10.1016/j.jmb.2017.09.009>.
- [81] N.M. Mascarenhas, V.L. Terse, S. Gosavi, Intrinsic disorder in a well-folded globular protein, *J. Phys. Chem. B* 122 (2018) 1876–1884, <https://doi.org/10.1021/acs.jpcc.7b12546>.
- [82] E. Haas, M. Wilchek, E. Katchalski-Katzir, I.Z. Steinberg, Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer, *Proc. Natl. Acad. Sci. U. S. A.* 72 (1975) 1807–1811, <https://doi.org/10.1073/pnas.72.5.1807>.
- [83] J.R. Alcalá, E. Gratton, F.G. Prendergast, Interpretation of fluorescence decays in proteins using continuous lifetime distributions, *Biophys. J.* 51 (1987) 925–936, [https://doi.org/10.1016/S0006-3495\(87\)83420-3](https://doi.org/10.1016/S0006-3495(87)83420-3).

- [84] E. Ben Ishay, G. Hazan, G. Rahamim, D. Amir, E. Haas, An instrument for fast acquisition of fluorescence decay curves at picosecond resolution designed for double kinetics experiments: application to fluorescence resonance excitation energy transfer study of protein folding, *Rev. Sci. Instrum.* 83 (2012) <https://doi.org/10.1063/1.4737632>.
- [85] M.P. Lillo, B.K. Szpikowska, M.T. Mas, J.D. Sutin, J.M. Beechem, Real-time measurement of multiple intramolecular distances during protein folding reactions: a multisite stopped-flow fluorescence energy-transfer study of yeast phosphoglycerate kinase, *Biochemistry.* 36 (1997) 11273–11281, <https://doi.org/10.1021/bi970789z>.
- [86] K.K. Sinha, J.B. Udgaonkar, Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: evidence for a continuous transition, *J. Mol. Biol.* 353 (2005) 704–718, <https://doi.org/10.1016/j.jmb.2005.08.056>.
- [87] G. Fuertes, N. Banterle, K.M. Ruff, A. Chowdhury, R. V. Pappu, D.I. Svergun, E.A. Lemke, Comment on “Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water”, *Science.* 361 (2018) eaau8230. doi:<https://doi.org/10.1126/science.aau8230>.
- [88] J.A. Riback, M.A. Bowman, A.M. Zmyslowski, K.W. Plaxco, P.L. Clark, T.R. Sosnick, Commonly used FRET fluorophores promote collapse of an otherwise disordered protein, *Proc. Natl. Acad. Sci.* 116 (2019) 8889–8894, <https://doi.org/10.1073/pnas.1813038116>.
- [89] P.G. de Gennes, Collapse of a polymer chain in poor solvents, *J. Phys. Lettres.* 36 (1975) 55–57, <https://doi.org/10.1051/jphyslet:0197500360305500>.
- [90] C. Williams, F. Brochard, H.L. Frisch, Polymer collapse, *Annu. Rev. Phys. Chem.* 32 (1981) 433–451, <https://doi.org/10.1146/annurev.pc.32.100181.002245>.
- [91] A. Halperin, P. Goldbart, Early stages of homopolymer collapse, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 61 (2000) 565–573, <https://doi.org/10.1103/PhysRevE.61.565>.
- [92] E. Kussell, E.I. Shakhnovich, Glassy dynamics of side-chain ordering in a simple model of protein folding, *Phys. Rev. Lett.* 89 (2002) 14–17, <https://doi.org/10.1103/PhysRevLett.89.168101>.
- [93] K. Weber, R.L. Jack, V.S. Pande, Emergence of glass-like behavior in Markov state models of protein folding dynamics, *J. Am. Chem. Soc.* 135 (2013) 5501–5504, <https://doi.org/10.1021/ja4002663>.
- [94] A.S.J.S. Mey, P.L. Geissler, J.P. Garrahan, Rare-event trajectory ensemble analysis reveals metastable dynamical phases in lattice proteins, *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 89 (2014) 1–8. doi:<https://doi.org/10.1103/PhysRevE.89.032109>.
- [95] E. Pitard, H. Orland, Dynamics of the swelling or collapse of a homopolymer, *Europhys. Lett.* 41 (1998) 467–472, <https://doi.org/10.1209/epl/i1998-00175-8>.
- [96] S. Mukhopadhyay, R. Krishnan, E.A. Lemke, S. Lindquist, A.A. Deniz, A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures, *Proc. Natl. Acad. Sci.* 104 (2007) 2649–2654, <https://doi.org/10.1073/pnas.0611503104>.
- [97] V.A. Voelz, M. Jäger, S. Yao, Y. Chen, L. Zhu, A. Steven, G. R. Bowman, M. Friedrichs, O. Bakajin, L.J. Lapidus, S. Weiss, V.S. Pande, Slow unfolded-state structuring in ACBP folding revealed by simulation and experiment supplementary figures and legends, *J. Am. Chem. Soc.* 134 (2012) 12565–12577.
- [98] H.S. Chan, K.A. Dill, Transition states and folding dynamics of proteins and heteropolymers, *J. Chem. Phys.* 100 (1994) 9238–9257, <https://doi.org/10.1063/1.466677>.
- [99] K.M. Hamadani, S. Weiss, Nonequilibrium single molecule protein folding in a coaxial mixer, *Biophys. J.* 95 (2008) 352–365, <https://doi.org/10.1529/biophysj.107.127431>.
- [100] B. Wunderlich, D. Nettels, S. Benke, J. Clark, S. Weidner, H. Hofmann, S.H. Pfeil, B. Schuler, Microfluidic mixer designed for performing single-molecule kinetics with confocal detection on timescales from milliseconds to minutes, *Nat. Protoc.* 8 (2013) 1459–1474, <https://doi.org/10.1038/nprot.2013.082>.
- [101] N. Aghera, J.B. Udgaonkar, Stepwise assembly of β -sheet structure during the folding of an SH3 domain revealed by a pulsed hydrogen exchange mass spectrometry study, *Biochemistry.* 56 (2017) 3754–3769, <https://doi.org/10.1021/acs.biochem.7b00374>.