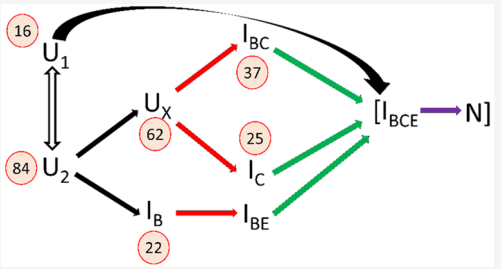
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Mapping Distinct Sequences of Structure Formation Differentiating Multiple Folding Pathways of a Small Protein

Sandhya Bhatia, Guruswamy Krishnamoorthy, and Jayant B. Udgaonkar*

Cite This: J. Am. Chem. Soc. 2021, 143, 1447−1457
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: To determine experimentally how the multiple folding pathways of a protein differ, in the order in which the structural parts are assembled, has been a long-standing challenge. To resolve whether structure formation during folding can progress in multiple ways, the complex folding landscape of monellin has been characterized, structurally and temporally, using the multisite time-resolved FRET methodology. After an initial heterogeneous polypeptide chain collapse, structure formation proceeds on parallel pathways. Kinetic analysis of the population evolution data across various protein segments provides a clear structural distinction between the parallel pathways. The analysis leads to a phenomenological model that describes how and when discrete segments acquire structure independently of each other in different subensembles of protein molecules. When averaged over all molecules, structure formation is seen to progress as α-helix formation, followed by core consolidation, then β-sheet formation, and last end-to-end distance compaction. Parts of the protein that are closer in the primary sequence acquire structure before parts separated by longer sequence.



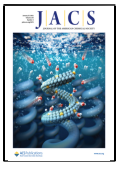
INTRODUCTION

The protein folding reaction is a "disorder to order transition" from the unstructured random-coil-like unfolded (U) state to the uniquely structured native (N) state.[1,2] Many experimental, computational, and theoretical studies have established the existence of multiple intermediates and heterogeneity in the folding of even small single-domain proteins.[3−5] A fundamental unanswered question concerns the uniqueness of the protein folding pathway.[6−9] Can a protein be assembled from its parts in more than one way? Do individual folding pathways of a protein describe distinct sequences of formation of progressively more structured conformations, or is there a defined pathway for hierarchical structure formation shared, with relatively minor distinctions, by different folding pathways?[1] The traditional view of protein folding is that a single unique sequence of structural events leads to the formation of the N state.[10−14] In contrast, energy landscape theory suggests that folding can occur via a multitude of paths spanning ensembles of intermediates and transition states that differ in structure and energy,[6,8,9,15,16] implying that a single unique sequence of folding events is very unlikely.[17,18] The existence of more than one folding pathway would be favored by evolution, as alternative pathways offer robustness to the process of folding[1,18,19] under varying environmental conditions, as expected *in vivo*.[20−23] It appears that the folding process is plastic enough to adjust to environmental changes, most probably by utilizing alternative routes to the N state.

Indeed, multiple folding pathways have been shown to be operative concurrently during the folding of several proteins.[1,24−31] Importantly, folding and unfolding have been observed to switch from one pathway to another when the conditions have been varied.[32−38] Moreover, in different folding/unfolding conditions,[25,35−39] or upon mutation,[32,34,40,41] the sequence of structural events can be different, as can be the cooperativity of the folding/unfolding reaction[4,42,43] and the structure present in a folding intermediate.[44,45] Single-molecule methods have also revealed the existence of multiple folding and unfolding pathways.[46−51] Although evidence supporting the existence of more than one folding/unfolding pathway continues to grow, there is little evidence for a large multitude of pathways as envisaged by energy landscape theory.

Measurements of time-resolved fluorescence decay kinetics of a fluorophore as a function of folding time[45,52−54] provide quantitative temporal and structural information about the populations of different conformations formed at different stages of folding. When such measurements are of site-specific Forster resonance energy transfer (FRET) and are analyzed by the maximum entropy method (MEM), distance distributions across different structural parts of the protein can be measured as a function of (un)folding reaction.[53,55−58] Multisite time-resolved FRET (trFRET) measurements, in conjunction with

Received: October 21, 2020
Published: January 12, 2021

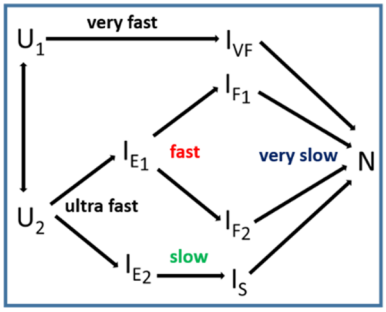


https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

MEM analysis, have been used extensively to characterize the heterogeneity of protein folding reactions in both equilibrium[56,59,60] and kinetic[45,54,58] studies of folding and unfolding. Such an experimental approach is expected to not only identify multiple folding pathways but also define how these pathways differ in their sequence of structure formation events.

The small monomeric protein MNEI is a variant of naturally occurring heterodimeric monellin, in which the C-terminus of chain B of the latter is covalently linked via a Gly-Phe dipeptide[29] to the N-terminus of chain A. MNEI has been used extensively as a model protein for protein folding studies. Hydrogen-exchange (HX-MS) studies have identified the sequence of structural events during folding and unfolding, ensemble-averaged at each time point of the reaction,[42] and have also revealed how folding cooperativity is determined by stability.[4] Multisite trFRET equilibrium studies of unfolding have also revealed the heterogeneity of unfolding, its noncooperative nature, and that the degree of noncooperativity differs between the lone helix and different parts of the β-sheet.[60] Similarly monitored kinetic studies of unfolding have identified two major pathways of unfolding and showed that both continuous and barrier-limited steps lead to the formation of the U state.[58] Pertinent to the current study, previous fluorescence-monitored measurements have identified ultrafast (>1000 s^{-1}), very fast (100 s^{-1}), fast 10 (s^{-1}), slow (0.1 s^{-1}), and very slow (0.001 s^{-1}) kinetic phases of folding.[29] Double-jump, interrupted folding experiments have shown that the very slow phase of folding leads to the formation of the N state and/or a state with N-like stability and that the four faster kinetic phases lead to the formation of intermediates on competing folding pathways. The very fast phase of folding originates from a subpopulation of molecules in the U state, which differ from the remaining molecules in the *cis*−*trans* isomerization status of peptidyl−prolyl bonds. The proposed kinetic mechanism (Scheme 1) has folding intermediates

Scheme 1



forming in fast and slow kinetic phases from the very early intermediates, whose structural heterogeneity has been confirmed by pulsed thiol labeling[30] and steady-state FRET[61] experiments, as well as by microsecond mixing experiments.[62]

While Scheme 1 accounts for the kinetic data, there is as yet no understanding of how the different intermediates differ in their structures and, hence, of how the different folding pathways differ in their sequences of structural events. A recent single-site trFRET study[53] has confirmed an important feature of the mechanism (Scheme 1) that initial polypeptide chain collapse leads to the formation of the two very early intermediate ensembles, whose relative populations depends on how stabilizing are the folding conditions. That study also suggested that the use of multiple FRET pairs in a trFRET study might not only reveal the full complexity of the folding mechanism but also reveal how the multiple folding pathways (Scheme 1) differ in their sequences of structural events.

In the current study, the folding of MNEI has been studied by monitoring FRET using four FRET pairs in different mutant variants (Figures 1 and S1). The FRET pairs were placed so that subpopulations of molecules that differed in the extent of separation of the donor from the acceptor could be distinguished on the basis of their distinct fluorescence lifetime and distinct distance distributions. It therefore became possible to determine how subpopulations of molecules that differed in a distance within the helix (H), a distance separating two strands in the β-sheet (B), a distance in the core (C) of the protein, and an end-to-end distance (E) evolved with time of folding. It is shown that changes in H, B, C, and E occur on multiple folding pathways that differ in the temporal order of the structural changes.

RESULTS AND DISCUSSION

Time-resolved fluorescence decay curves (Figure S2) were measured for four single Trp, single Cys containing mutant variants of MNEI, and their thionitrobenzoate (TNB)-labeled counterparts, as a function of the time of folding in 0.4 M GdnHCl (Figure 1). For each pair of unlabeled and labeled variants, the mean fluorescence lifetimes (averaged over every 100 ms) were determined at 100 ms intervals during folding (Figure S3). The mean fluorescence lifetime is the amplitude-weighted average of the individual lifetime components obtained from a multiexponential fit to the fluorescence decay traces ($\tau_m = \sum \alpha_i \tau_i$; $\sum a_i = 1$; see Materials and Methods section, SI, for details). The FRET efficiency was calculated at each time point, and a kinetic trace was obtained that monitored folding at each of the four segments, with ensemble-averaged FRET efficiency as the probe (Figure 1). Such trFRET-monitored kinetic traces were obtained for segments H (using W19C29 and W19C29-TNB), B (using W4C42 and W4C42-TNB), C (using W19C42 and W19C42-TNB), and E (using W4C97 and W4C97-TNB) (see above). Each kinetic trace showed the FRET efficiency changing in a burst phase (100 ms), a fast kinetic phase, and a slow kinetic phase. A very slow change in FRET efficiency could not be monitored. Global fitting showed that the same value for the fast folding rate constant and for the slow folding rate constant could describe well all four kinetic traces (Figure S4).

Importantly, the global fitting also showed that the apparent rate constants determined for the fast and slow kinetic phases of trFRET-monitored folding also described the kinetic phases of folding monitored by far-UV CD measurement (Figure S4). This result and a comparison of the kinetics of folding of labeled and unlabeled proteins (Figure S5) indicated that the mutations and the introduction of the FRET label had no significant effect on the folding rate constants. Earlier studies had also shown that the structure and stability[58,60,61] were only minimally affected by such mutation and addition of the label. Hence, the differences observed between the four FRET-monitored kinetic traces (Figure 1) in the relative amplitudes of the fast and slow phases of folding (Figure S4) cannot be attributed to structural perturbations caused by mutation and/or labeling.

1. FRET-Monitored Folding Kinetics Are Different at Different Intramolecular Segments. The observation that the ensemble-averaged trFRET-monitored kinetic traces for

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

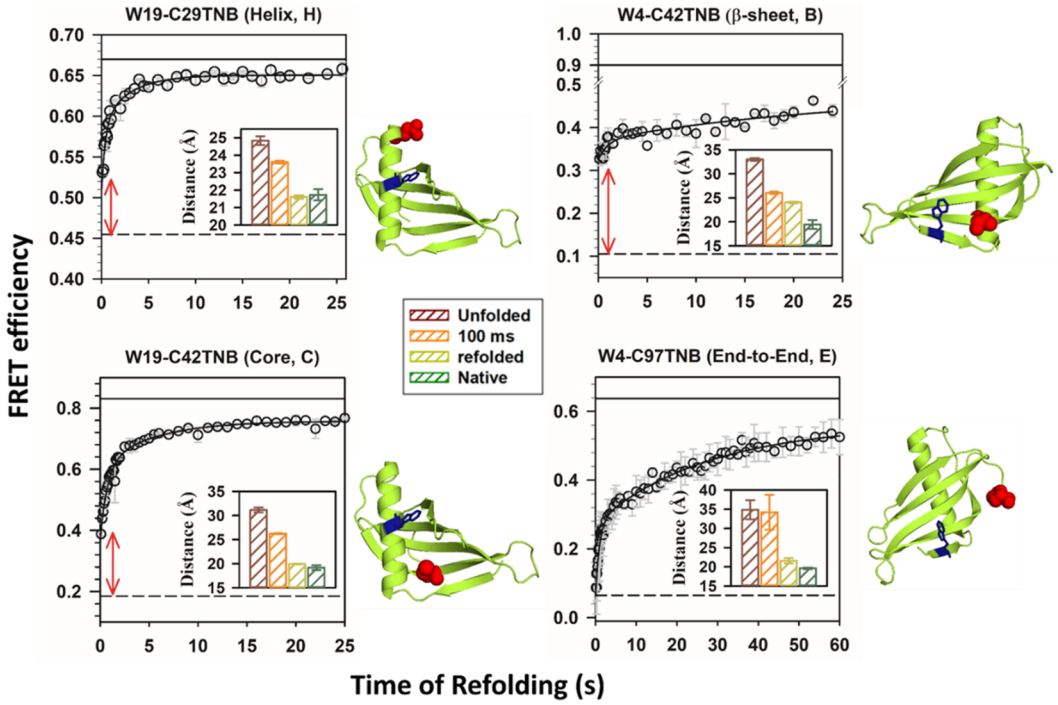


Figure 1. Multisite trFRET-monitored kinetics delineates segment-specific differences in collapse and folding. Kinetics of folding of MNEI in 0.4 M GdnHCl at pH 8, 25 °C, monitored by multisite trFRET. The different panels correspond to data for different intramolecular FRET pairs, as indicated on the top of each panel. The positions of the donor (Trp, shown as blue sticks) and acceptor (Cys-TNB, shown as red spheres) fluorophores of each FRET pair are shown in the structure on the right of each panel. The structures (pdb ID 1IV9) were drawn using Pymol. The FRET efficiency, in each case, was determined by using the mean lifetime values for the unlabeled and corresponding TNB-labeled mutant variant, at each folding time. In each panel, the solid and dashed black horizontal lines represent the FRET efficiency in the N state (in 0.4 M GdnHCl) and the U state (in 4 M GdnHCl), respectively. In each panel, the red line with arrowheads on both ends corresponds to the burst phase change in FRET efficiency between the unfolded state and at the first observable time (100 ms) of folding. The inset in each panel shows the average donor−acceptor pair distance ($<R_{DA}>$) determined at different times of folding. The distances were determined using the FRET efficiency values and the Forster distance values in the Forster equation.[60] The error bars represent the standard errors of measurements from two independent double kinetics experiments. The kinetic parameters obtained from global fitting of the data (fits shown as solid lines through the kinetic data) are shown in Figure S4.

folding at the four sequence segments differed not in the rate constants but in the relative amplitudes of the fast and slow phases (Figures S4 and S5) can be interpreted in two ways: (1) the change in the FRET-monitored segmental distance occurs in two sequential steps, and hence, the relative amplitude of each of the two phases is the fractional change in distance on going from the initial state to the intermediate state and then from the intermediate state to the final state; (2) the change in the FRET-monitored segmental distance occurs along two parallel pathways; that is, there are two subpopulations of molecules with U-like distances, which encounter different energy barriers to undergoing further collapse and folding.

The burst phase increase in FRET efficiency, observed for the H, B, and C segments, indicated that at 100 ms, the first observable time of folding in the current study, the intramolecular distances monitoring these segments were significantly shorter than in the U state (insets, Figure 1). It should be noted that these distances were determined directly from the mean lifetimes using the Forster equation (see SI Materials and Methods) and should therefore be considered as rough and qualitative measures of compaction. When these distances were calculated assuming that they have Gaussian distributions, they were found to have values slightly different from those reported in the insets of Figure 1, but the trends of compaction were found to be the same (data not shown; see SI Materials and Methods, for the limitation of this assumption). This burst phase (<100 ms) reduction in the segmental distances leads to the formation of a collapsed ensemble, which is known to precede structure formation in the case of monellin.[53,61,62] The extent of collapse was different for the different segments in the collapsed ensemble, indicative of an overall asymmetric initial polypeptide chain collapse.

2. The Collapsed Ensemble Consists of Expanded and Collapsed Subpopulations. In a previous study,[53] it had been observed that the fluorescence lifetime distribution observed for the collapsed ensemble at 100 ms of W19C42-TNB, in which the C segment was monitored, was bimodal. The peak at shorter lifetimes (<0.6 ns) was N-like, as it was centered near the distribution observed for the N state, and the peak at longer lifetimes (>0.6 ns) was U-like, as it was centered near the distribution of the U state. From the bimodal distribution, it was possible to determine the fraction of molecules in the collapsed ensemble that were U-like and N-like at segment C. Figure 2 shows that the fluorescence lifetime distributions observed at 100 ms of folding for W4C42-TNB and W4C97-TNB were also bimodal, and it was therefore possible to determine, from the relative sum of amplitudes (see legend to Figure 3), the fractions of molecules in the collapsed ensemble at 100 ms of folding that were U-like at segments B, C, and E (Figure 3A).

In order to correctly determine the fractions of molecules that were N-like and U-like at a particular segment in the

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

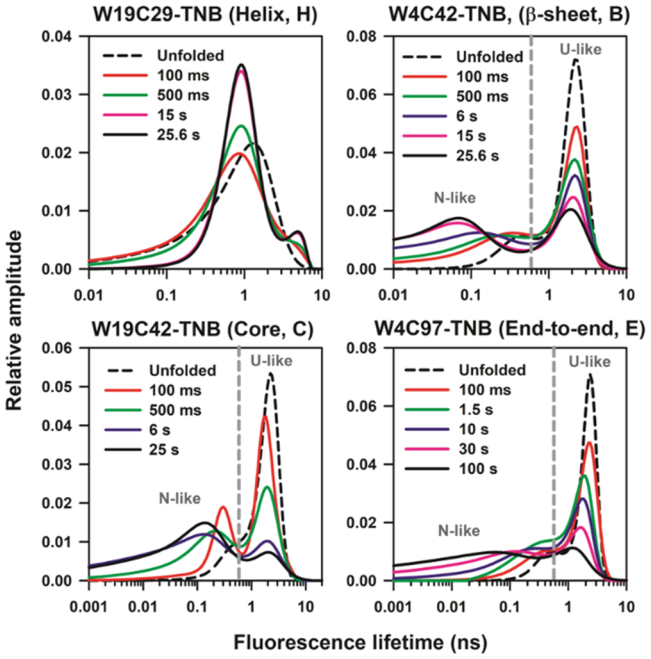


Figure 2. Evolution of fluorescence lifetime distributions as a function of folding time. MEM analysis-derived fluorescence lifetime distributions were obtained for the different TNB-labeled variants at various times of folding at pH 8, 25 °C. The four panels correspond to different TNB-labeled protein variants that report on changes in different structural regions of the protein (as indicated on the top of each panel). The solid and dashed black curves in each panel indicate the lifetime distributions for the refolded and unfolded states, respectively. Distributions corresponding to different times of the folding reaction are shown in different colors as described in each panel. In each panel, the U-like and N-like labels correspond to the fluorescence lifetime distributions peaked at longer (>0.6 ns) and shorter (<0.6 ns) lifetimes, respectively. The gray vertical dashed line (at 0.6 ns) divides the distributions between the U-like and N-like subensembles. The *x*-axis has been plotted on a log-scale. The relative amplitude on the *y*-axis has been normalized to the sum of amplitudes for each distribution to make the total population fraction equal to 1. The fraction of molecules having a U-like or N-like distance for a particular segment is given by the relative sum of amplitudes for the corresponding distribution.

collapsed ensemble at 100 ms or, indeed, at any subsequent time of folding, it was important to take into account the observation that for any segment the equilibrium U state had a fraction of molecules that had an N-like distribution (<0.6 ns) and that the equilibrium N state had a fraction of molecules that had a U-like distribution (>0.6 ns) (Figure 2). These fractions were similar for each pair of unlabeled (Figure S6) and corresponding labeled (Figure 2) unfolded proteins, indicating that they originate from different electronic structures of the fluorophore or from different Trp rotamers[63,64] and not from the presence of the quenching TNB moiety. With Y_U equal to the relative sum of amplitudes for the U-like distribution observed for the equilibrium U state, Y_N equal to the relative sum of amplitudes for the U-like distribution observed for the equilibrium N state, and Y_i equal to the relative sum of amplitudes for the U-like distribution observed at any time of folding, the fraction of molecules (f_U) expanded (U-like) at a particular segment could be determined as $f_U = (Y_i - Y_N)/(Y_U - Y_N)$. It should be noted that f_U was determined from MEM analysis-derived fluorescence lifetime distributions, such as those in Figure 2, because such distributions have been shown previously to accurately estimate the relative fractions of N-like and U-like molecules present together.[53,55] f_U was not determined from discrete analysis (see SI Materials and Methods) of the fluorescence decay curves, because the values of the rate constants and relative amplitudes obtained from discrete multiexponential fits, which assume two to four discrete states, may not be accurate. It is well known that different combinations of relative amplitudes and rate constants can yield equally good fits even for two-exponential fits.[65] Hence, it is not possible to confidently determine the fractions of molecules that were N-like and U-like at any segment from the relative amplitudes obtained from discrete analysis, as is also apparent in Figure S2.

Different fractions of molecules in the collapsed ensemble at 100 ms of folding were N-like at the three segments (Figure 3A): 16% of the molecules had collapsed to N-like dimensions at segment E, 23% at segment C, and 45% at segment B. The collapsed ensemble formed at 100 ms of folding was clearly heterogeneous with different fractions of molecules collapsed at one or more segments or at none of the segments. For three segments, with each being either expanded or collapsed, the ensemble at 100 ms was expected to consist of eight subensembles. It was not necessary that the populations of these eight subensembles, which would be defined by their stabilities, be the same. It was therefore useful to assume minimal heterogeneity in the ensemble and examine the data in Figure 3A based on this assumption. Based on the previous observation[29,61] that the collapsed globule is a kinetic molten globule and on the expectation that molten globule-like molecules are likely to have multiple collapsed segments,[66] it was assumed that 16% of the molecules were present in an I_{BCE} subensemble in which segments B, C, and E had all collapsed to become N-like, 7% were present in an I_{BC} subensemble in which only segments B and C had collapsed to become N-like, 22% were present in an I_B subensemble in which only segment B had collapsed to become N-like, and 55% of the molecules had not collapsed but remained U-like (U_X) at all three segments (Figure 3B). The I_{BCE} subensemble corresponds to the molten globule-like I_{MG} subensemble identified in an earlier single-site trFRET study.[53] The assumptions therefore led to a description of the collapsed ensemble at 100 ms as being composed predominantly of four subensembles, each of which was populated to greater than 5%. The other four expected subensembles would also be populated, but appear to be too unstable to be populated to an extent sufficient to contribute significantly to subsequent folding. It is important to note here that the assumption that the collapsed globule has molecules with segment B collapsed, molecules with segments B and C collapsed, and molecules with segments B, C, and E collapsed does not in any way define any pathway of segments collapsing during the initial collapse process. Indeed, initial chain collapse, during the folding of MNEI, has been shown to occur via more than one pathway.[29,30] It is also not necessary that subsequent folding from the collapsed globule utilize the same pathways of segments becoming N-like, especially since the starting state (U) for initial chain collapse is different from the starting states for subsequent folding.

Interestingly, the fluorescence lifetime distribution seen at 100 ms of folding for W19C29-TNB, which reports on segment H, was unimodal, indicating the absence of population-level heterogeneity for this segment. This suggested that the helical structural content was the same in all molecules

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

A

Monitored structural segment ↓	Fraction (un)folded at the monitored segment at 100 ms of folding	
	U-like (unfolded)	N-like (folded)
β-sheet, B (W4C42-TNB)	0.55 ± 0.07 *(55)	0.45 ± 0.07 *(16+7+22)
Core, C (W19C42-TNB)	0.77 ± 0.03 *(55+22)	0.23 ± 0.03 *(16+7)
End-to-end, E (W4C97-TNB)	0.84 ± 0.08 *(55+22+7)	0.16 ± 0.08 *(16)

* The values in brackets are the simplest possible combination of sub-populations of molecules, which predicts the fractions of molecules that are U-like and N-like at a monitored segment (see text and panel B).

B

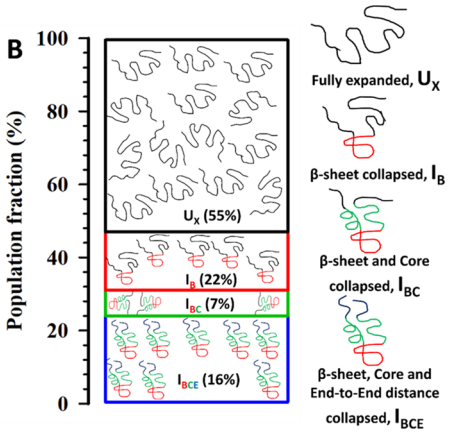


Figure 3. Resolving heterogeneity in the initial collapse transition. (A) Fraction of molecules, in the U-like (unfolded) and N-like (folded) distributions at the monitored structural segment, measured from relative sum of amplitudes for the corresponding peak (see Figure 2) at 100 ms of folding reaction. The relative sum of amplitudes is the sum of amplitudes of the distributions for the U-like or N-like distances divided by the sum of amplitudes of the distributions for both the U-like and N-like distances. The fractional population of molecules that are expanded (U-like) at a particular segment (designated as fraction unfolded, f_U)[53] was determined relative to the equilibrium N ($f_U = 0$) and U ($f_U = 1$) states [$f_U = (Y_i - Y_N)/(Y_U - Y_N)$; Y_i is the relative sum of amplitudes for the U-like distance distribution at any given time of folding] (Figure S8). (B) Schematic describing the minimal heterogeneity observed in the collapsed ensemble at 100 ms of folding, which explains the data shown in panel A. The polymer cartoons represent conformationally different subensembles with either none, one, two, or three distances collapsed to native-like dimensions. The black segments are expanded. The red, green, and blue segments are collapsed in the β-sheet, core, and the end-to-end distance, respectively. The coexistence of U_X (fully expanded), I_B (β-sheet collapsed), I_{BC} (β-sheet and core collapsed), and I_{BCE} (fully collapsed) highlights the underlying heterogeneity in the collapsed ensemble at 100 ms of folding.

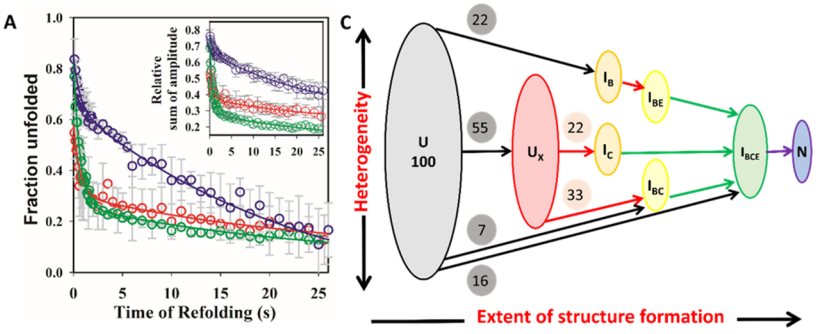
at 100 ms of folding, whether the molecules belonged to the U_X, I_B, I_{BC}, or I_{BCE} subensemble. Interestingly, the fluorescence lifetime distribution remained unimodal during the course of folding (Figure 2, W19C29-TNB). In the case of unlabeled W19C29, a significant decrease in the exposure of Trp19 to polar solvent molecules during folding resulted in a gradual increase in the fluorescence lifetime (Figures S3, S6, and S7). However, for W19C29-TNB, the extent of increase in the fluorescence lifetime of W19 during folding was reduced significantly as compared to unlabeled W19C29, as evident from both discrete (Figure S3) and MEM analysis (Figure S7) of the data. This observation highlights the role of FRET in reducing the lifetime of the native state. Quantification of the peak movement as a function of folding time (Figure S7), after translating the fluorescence lifetime data for W19C29 and W19C29-TNB into intrasegmental distance,[53,60] suggested that segment H contracted continuously in all protein molecules, whether in U_X, I_B, I_{BC}, or I_{BCE}, as the folding reaction progressed.

3. Segmentwise Population Evolution Kinetics of Different Parts of the Protein Provides a Sequence of Structure Formation along Parallel Pathways. The time evolution of the MEM analysis-derived fluorescence lifetime distributions was used to extract the fraction of molecules that were expanded (U-like) at each of the three segments (B, C, or E) as a function of folding time (Figures 2, S8, and 4A). For each segment, the fraction of molecules that were expanded (U-like) at that segment decreased in two observable kinetic phases, a fast phase and a slow phase. For all the segments, the rate constants for the fast phase were found to be similar, and so were the rate constants of the slow phase (Figure S5). This result was surprising, as it indicated that the barrier heights to collapse of the different segments were similar, but it also indicated that if collapse of the segments occurred on multiple pathways with fast and slow phases, there would be flux of folding molecules along all pathways. Unfortunately, the individual rate constants were too similar in value, with the measured differences being within the errors of measurement (Figure S5), to deduce any small variation in barrier heights on different pathways. In fact the rate constants were similar enough that it was possible to carry out a global fit of the kinetic data for all the segments. The same rate constant for the fast phase and the same rate constant for the slow phase described well the folding at all three segments (Figures 4B and S9).

The two kinetic phases observed for the decrease in the population of the U-like segments in the different subensembles correspond, in terms of their rate constants, to the fast and slow kinetic phases of folding observed in earlier fluorescence-monitored ensemble-averaging kinetic measurements.[29,30,61] These rate constants decrease with an increase in the GdnHCl concentration, indicating that both kinetic phases correspond to folding reactions.[29] The possibility that either of the two kinetic phases observed in the current trFRET study corresponds to an unfolding reaction that serves as an off-pathway correction of an optional error[67] present in the fold of the subensembles possessing one or more U-like segments can, therefore, be ruled out. It should also be noted that there is no evidence for any backtracking during the folding reaction.[54]

There were two important observations that had to be taken into account for deducing a model to describe folding. (1) The collapsed ensemble at 100 ms of folding was heterogeneous, consisting of subensembles of molecules that were expanded (U-like) at only segment E (I_{BC}), at only segments C and E (I_B), and at segments B, C, and E (U_X), and that there was also a subensemble that was not expanded but compact (N-like) at all three segments (I_{BCE}) (Figure 3). (2) The fraction of molecules transforming from being U-like to becoming N-like at a segment was different for each segment (Figure 4A and B), indicating that multiple pathways of folding were operative. A fraction of molecules became N-like at a segment (B, C, or E) in the fast kinetic phase on one or more pathways and in the

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457



B

Monitored structural segment (sub-ensembles in which the segment is expanded) ↓	Ensemble at 100 ms: Fraction unfolded (%) at the monitored segment ↓	Subpopulation types →	U_X (55)	I_B (22)	I_{BC} (7)	I_{BCE} (16)
		Kinetics of disappearance of expanded (U-like) form				
		Fast phase 1.56 ± 0.01 s^{-1}		Slow phase 0.04 ± 0.005 s^{-1}		
		Relative amplitude		Relative amplitude		
		Observed	**Predicted	Observed	**Predicted	
β-sheet, B (U_X)	55 ± 7	0.60 ± 0.13 *(33)	0.60 (= 33/(22+33)) $U_X \rightarrow I_{BC}$	0.40 ± 0.13 *(22)	0.40 (= 22/(22+33)) $I_C \rightarrow I_{BCE}$	
Core, C ($U_X + I_B$)	77 ± 3	0.73 ± 0.04 *(56)	0.71 (= (22+33)/(22+22+33)) $U_X \rightarrow I_C$ & $U_X \rightarrow I_{BC}$	0.27 ± 0.04 *(21)	0.29 (= 22/(22+22+33)) $I_{BE} \rightarrow I_{BCE}$	
End-to-End, E ($U_X + I_B + I_{BC}$)	84 ± 8	0.20 ± 0.08 *(17)	0.26 (= 22/(22+22+33+7)) $I_B \rightarrow I_{BE}$	0.80 ± 0.08 *(67)	0.74 (= (22+33+7)/(22+22+33+7)) $I_C \rightarrow I_{BCE}$ & $I_{BC} \rightarrow I_{BCE}$	

Figure 4. Sequence of structure formation obtained from the population-evolution kinetics derived from MEM analysis. (A) Kinetics of conversion of expanded (U-like) distances to the collapsed (N-like) distances in the different subensembles, as a function of folding time. The differently colored circles correspond to fractions of molecules unfolded at different segments: red, B; green, C; blue, E. The fraction of molecules that are U-like (unfolded) and N-like (folded) at the monitored structural segment was measured from the relative sum of amplitudes for the corresponding peak (inset, panel A). The relative sum of amplitudes and the fraction unfolded, f_U, were determined as described in the legend to Figure 3. The error bars represent the standard errors of measurements from two independent double kinetics experiments. The observed kinetics were fitted globally to the sum of two exponentials, with the same two rate constants but variable amplitudes for all three FRET pair variants. The rate constants and the relative amplitudes obtained for various FRET pairs from the global fit are given in the table in panel B. (C) Phenomenological model describing the evolution of structural heterogeneity as folding progresses from the U to the N state. The scheme predicts the relative amplitude data in panels A and B. The extent of structure formation increases along the *x*-axis. Conformational entropy (heterogeneity) decreases with folding, as depicted in the width of the shown scheme. 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. The numbers in the circles denote the percentages of molecules following a given folding route.

slow kinetic phase on other, mutually exclusive, pathways. If a single sequential pathway led from the expanded (U-like at any structural segment) to the collapsed (N-like) populations of molecules, the population of U-like molecules would have decreased in a single kinetic phase. A previous study[53] had shown that the invocation of at least two parallel pathways leads to a very good fit of the population kinetics data to a quantitative coarse-grained Markov evolution model of the folding of MNEI.

A phenomenological model for folding (Figure 4C) could be deduced from the data (Figure 4B). (1) At 100 ms of folding, 77% of the molecules were expanded at segment C, as U_X and I_B (Figure 3); 73% of these 77% molecules, that is, 56% of all molecules, became compact at segment C in the fast phase of folding (Figure 4A and B). Since 55% of the molecules present at 100 ms were U_X (Figure 3), the simplest deduction was that all U_X molecules became compact at segment C in the fast phase. (2) At 100 ms, only the 55% of molecules present as U_X were expanded at segment B (Figure 3); hence, the fast phase of compaction at segment B had necessarily to begin only from U_X; 60% of these 55% molecules, that is, 33% of all molecules, became compact at segment B in the fast phase (Figure 4A and B). (3) Hence, it was deduced that during the fast phase of folding starting from U_X 33% of all molecules became compact at both segments B and C to form I_{BC}, and 22% became compact only at segment C to form I_C. (4) 84% of the molecules present at 100 ms as U_X, I_B, and I_{BC} were expanded at segment E (Figure 3); 20% of these 84% molecules, that is, 17% of all molecules, became compact in the fast phase of folding (Figure 4A and B). Since 22% of all molecules present at 100 ms were I_B, it was deduced that all I_B molecules became compact also at segment E in the fast phase to form I_{BE}.(5) At the end of the fast phase of folding, only 22% of the molecules were expanded at segment B, as I_C. Hence, slow phase compaction of segment B had to start only from I_C. (6) At the end of the fast phase of folding, only 22% of the molecules were expanded at segment C, as I_{BE}. Hence, slow phase compaction at segment C had to start only from I_{BE}. (7) At the end of the fast phase, segment E was expanded in the 22% molecules present as I_C and in the 40% (33 + 7) of molecules present as I_{BC}. Hence, slow phase compaction of segment E had to start from the 62% of molecules present as I_C and I_{BC} at

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

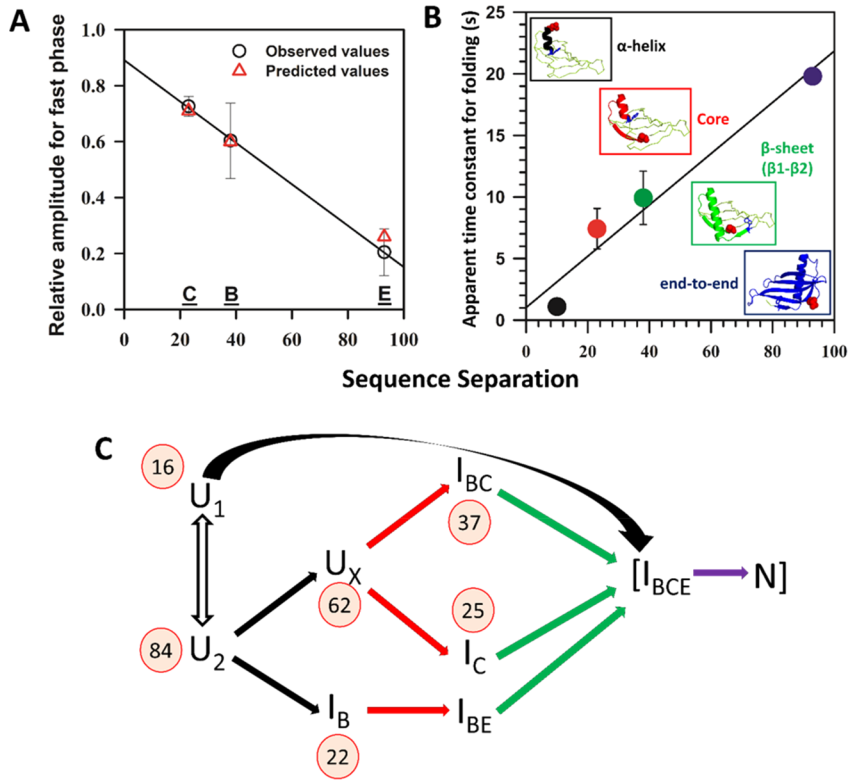


Figure 5. Sequence of structural events during the folding of MNEI. (A) Dependence of the relative amplitudes (observed and predicted from the scheme) for the fast kinetic phase of folding on sequence separation for the different FRET pairs. (B) Dependence of the apparent time constant on sequence separation for the different FRET pairs. The apparent time constant is determined as the amplitude-weighted average of the time constants ($\tau_{av} = \frac{\alpha_1\tau_1 + \alpha_2\tau_2}{\alpha_1 + \alpha_2}$; $\tau_i = \frac{1}{k_i}$) observed for the two exponentials in the population evolution kinetics (Figure 4B) for all the FRET pairs except for the α-helix. For the α-helix, the apparent time constant was determined from the unimodal contraction kinetics (Figure S4) which leads to the formation of the N state. The different structures shown inside panel B represent the overall sequence of structure formation obtained from the apparent time constants. Different structural segments are shown in different colors corresponding to the various FRET pairs; all the residues spanned between a given FRET pair are shown with one color. The positions of the FRET donor and acceptor are shown as a blue ring and as a red sphere, respectively. The color scheme is as follows: black, half-helix (segment H); red, core (segment C); green, β-sheet (segment B); and blue, end-to-end (segment E). The error bars represent the standard errors of measurements from two independent double kinetics experiments. The folding mechanism is shown in panel C. The black, red, green, and violet arrows represent the unobservable (over within 100 ms), fast, slow, and very slow kinetic phases of folding reaction, respectively. The U state ensemble (U_1 and U_2) gives rise to three subensembles: U_X, I_B, and I_{BCE}. I_{BCE} continues to evolve gradually to form the N state. Furthermore, I_{BCE} formation from U_X and I_B occurs via three independent parallel pathways, each consisting of two sequential cooperative steps, in fast and slow kinetic phases, as shown in the scheme. The numbers in the circles represent the percentages of molecules following a given folding route.

the end of the fast phase. (8) At the end of the slow phase all molecules had become compact at the B, C, and E segments to form I_{BCE}; 22% had formed from I_{BE}, 22% from I_C, and 40% from I_{BC} during the slow phase of folding, while the remaining 16% had formed directly from U.

The observed kinetics of compaction at the different segments (Figure 4A and B) can be seen to be consistent with the phenomenological model (Figure 4C). The model shows that 33% of molecules become N-like at segment B during the fast $U_X \rightarrow I_{BC}$ transition, and 22% of molecules become N-like during the slow $I_C \rightarrow I_{BCE}$ transition (Figure 4C). Hence, the model predicts relative amplitudes of 0.60 and 0.40 for the fast and slow phases of the decrease in the population of molecules that were U-like at segment B at 100 ms (Figure 4B). At segment C, 22% of the molecules become N-like during the fast $U_X \rightarrow I_C$ transition, 33% of molecules become N-like during the fast $U_X \rightarrow I_{BC}$ transition, and 22% of molecules become N-like during the slow $I_{BE} \rightarrow I_{BCE}$ transition (Figure 4C). Hence, the model predicts relative amplitudes of 0.71 and 0.29 for the fast and slow phases, respectively, of the decrease in the population of molecules that were U-like at segment C at 100 ms (Figure 4B). At segment E, 22% of molecules become N-like during the fast $I_B \rightarrow I_{BE}$ transition, 22% of molecules become N-like during the slow $I_C \rightarrow I_{BCE}$ transition, and 40% (33 + 7) of molecules become N-like during the slow $I_{BC} \rightarrow I_{BCE}$ transition (Figure 4C). Hence, the model predicts relative amplitudes of 0.26 and 0.74 for the fast and slow phases, respectively, of the decrease in the population of molecules that were U-like at segment E at 100 ms (Figure 4B). Thus, Figure 4B shows that the observed fractions of molecules that become N-like in the fast and slow kinetic phases, for each segment, are consistent with the phenomenological model.

The phenomenological model describing the folding of MNEI (Figure 4C) has features that are likely to be generally relevant to how several proteins fold. (1) There are multiple pathways and there are multiple steps on each pathway. Structural events occur in fast and slow phases on each pathway. (2) Structural events, such as core consolidation in segment C and β-sheet formation in segment B, can happen

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

independently of each other. (3) A particular structural event may happen in molecules in which other structural events may or may not have already occurred. A structural event may occur in the fast phase on one pathway and in the slow phase on another pathway. (4) The rate constant at which a structural event occurs depends on whether or not other structural events have already occurred in the same population of molecules. (5) The rate constants of structure formation along different pathways involving different regions are the same. This suggests that the underlying physicochemical interactions that act cooperatively during the different stages of folding are similar, irrespective of the structure-forming part. It should, however, be noted that the nature of the barriers that dictate the relative fluxes of molecules on the parallel pathways is yet to be understood. (6) Very importantly, these results clearly contradict the notion that structure accumulation during folding must occur in a unique manner and on a single defined pathway.[7,13,14]

4. Formation of Short-Range Local Contacts Precedes That of Long-Range Nonlocal Contacts. A strong linear correlation was observed between the relative amplitude of the fast kinetic phase and sequence separation (Figure 5A). In most of the molecules, the fast kinetic phase involves the formation of local short-range contacts such as in the core (segment C) and between the two strands in the β-sheet (segment B). It appears that if core consolidation occurs in the fast kinetic phase, then further reduction of the end-to-end distance occurs in the slow kinetic phase. Conversely, if the end-to-end distance becomes N-like in the fast kinetic phase, starting from the I_B molecules, then core consolidation occurs in the slow kinetic phase. This suggests that the consolidation of the core retards the end-to-end contact formation, and that if the end-to-end distance becomes N-like, it retards the core formation.

Despite the presence of the multiple pathways on which the same structural event may occur either fast or slow, it is possible to deduce the sequence of structure formation averaged over the entire population of molecules. Amplitude-averaged time constants (Figure 5B) obtained from the MEM analysis-derived population evolution kinetics for the different structural events (Figures 4A,B and 5A) and the time constant corresponding to single-exponential continuous contraction observed for the helix formation (Figure S7) indicate the most probable sequence of structural events from the U_X subensemble to the I_{BCE} subensemble. The relative flux of folding molecules down the four parallel pathways suggests that, on average, helix formation occurs first, followed by core consolidation, followed by β-sheet formation, and, finally, overall compaction of the end-to-end distance. These results agreed with the kinetics of the change in average distances (Figures S4 and S9). Clearly, the hierarchic accumulation of structure does not necessitate that folding should occur via a single defined pathway,[11,19,67,68] as has been suggested earlier.[7,13]

The same sequence of structure formation can be predicted from the contact order in the primary sequence,[69] as seen from the linear correlation between the amplitude-averaged (apparent) time constant and sequence separation (Figure 5B). A similar quantitative agreement has been observed for the dependence of the overall folding rate constant on relative contact order across multiple proteins earlier.[69] This is perhaps the first experimental demonstration of the linear dependence of intramolecular segmental folding rate constants on contact order, as predicted theoretically.[70,71] These results are also in accordance with classical theoretical predictions suggesting that short-range contacts play a predominant role in the folding kinetics and that long-range contacts determine the cooperativity and stability of the folded state.[70,71]

Importantly, the observed overall sequence of structure formation is in good accordance with the reverse of the sequence of unfolding events reported earlier in HX-MS studies.[42] It had been shown that the separation of the helix from the β-sheet during unfolding leaves an intact helix; this is similar to the result of the present study showing that the helix formation precedes the formation of the core, i.e., the helix−β-sheet interface. It had also been shown earlier that the $\beta2-\beta3$ strands unfolded as the last step. The current study also suggests that $\beta2$ (part of the core) forms early during folding.

5. Origin of Heterogeneity in the Initial Collapse Reaction. At 100 ms of folding, the ensemble of protein molecules consists of four intermediate subensembles of molecules: U_X (55%), I_B (22%), I_{BC} (7%), and I_{BCE} (16%) (Figure 3B). U_X is, however, likely to have formed within the first millisecond of folding,[53,62] and hence, both the $U_X \rightarrow I_{BC}$ reactions and $U_X \rightarrow I_C$ reactions (Figure 4C), whose time constant is about 1 s, would have progressed about 10% at 100 ms. It could therefore be possible that the ∼7% I_{BC} molecules present at 100 ms originate completely from U_X. This would suggest that the initial collapse of U, which is complete within 1 ms,[61,62] results in the formation of only three subpopulations of intermediates, U_X (62%), I_B (22%), and I_{BCE} (16%) (Figure 5C). The origin of this more limited heterogeneity at the end of the initial collapse reaction can possibly be understood by considering U state heterogeneity (Scheme 1) in terms of peptidyl−prolyl bond isomerization. Native MNEI has six Pro residues, of which Pro41 and Pro93 have *cis* peptidyl−prolyl bonds. The minor U_1 subensemble, which was shown[29] to comprise ∼15% of the U state ensemble (Scheme 1), also appears to have Pro41 and/or Pro93 in a *cis* conformation, as it folds in a very fast phase.[29] It is possible that the I_{BCE} molecules present at 100 ms arise from U_1 (Figure 5C). It is also possible that the isomerization status of peptidyl−prolyl bonds plays a role in the partitioning of the U_2 subensemble while folding to the U_X and I_B subensembles. The current understanding of the origin of the early heterogeneity is clearly far from satisfactory and will be the focus of future studies.

6. Folding to the Native State Occurs via Structurally Distinct Parallel Routes. Only a few experimental studies have been able to identify structural differences between parallel pathways for (un)folding. In the case of hen lysozyme, the α-domain and β-domain have been shown to fold via different pathways.[24] For other proteins, structural differences between transition states of (un)folding on parallel pathways have been discerned in some cases.[34−36,51]

Remarkably, the phenomenological model proposed in the current study (Figure 5C) is in broad agreement with the kinetic model proposed earlier[29,30] (Scheme 1). The present study has allowed the multiple pathways to be distinguished on the basis of the structural events that occur (Figure 4C). The major route from U_X to I_{BCE} involves sequential formation of local short-range contacts and then nonlocal long-range contacts, as expected from a hydrogen-bonding-driven hierarchical model of protein folding.[18] The minor route ($I_B \rightarrow I_{BCE}$) involves the reverse sequence of structure formation. As discussed earlier, early compaction of the end-to-end distance appears to result in the retardation of core

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

consolidation. Thus, the minor pathway of folding originating from I_B involves first the formation of relatively nonlocal contacts, followed by later consolidation of the core, which is dominated by local interactions. Another minor route that arises from the U_1 subpopulation ($U_1 \rightarrow I_{BCE}$) suggests that the isomerization status of a Pro residue plays a role in determining the sequence of structure formation during folding, by promoting the early formation of both long-range and short-range contacts simultaneously (and hence a very fast overall compaction to I_{BCE}).[72] In the future, this study will be extended to cover other segments of the protein to improve the structural resolution and to perhaps reveal even more complexity in the folding mechanism.

CONCLUSION

In the present study, the site-specific structural heterogeneity of the collapse and folding reactions of monellin has been resolved using multisite trFRET measurements. The collapsed ensemble present at 100 ms is shown to comprise different subensembles: a random-coil-like contracted metastable intermediate state (U_X), a site-specifically and nonuniformly collapsed intermediate (I_B), and a globular highly compact molten-globule intermediate (I_{BCE}). Subsequent folding occurs along multiple pathways differing in their sequence of events (Figures 4C and 5C). The helix undergoes collapse and folding in a completely barrierless gradual manner involving a continuous reduction in size, homogeneously across all molecules. When structural information for all molecules is averaged based upon the relative flux of folding molecules down the four parallel pathways, it appears that the sequence of events during folding is helix formation, followed by core consolidation, followed by the β-sheet formation, and finally end-to-end contact formation (Figure 5B). This averaged sequence of structure development seems to suggest that local structure consolidation precedes the formation of global structure.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.0c11097.

Materials and Methods section and supplementary figures (PDF)

AUTHOR INFORMATION

Corresponding Author

Jayant B. Udgaonkar − *National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560 065, India; Indian Institute of Science Education and Research, Pune 411 008, India;* orcid.org/0000-0002-7005-224X; Email: jayant@ncbs.res.in, jayant@iiserpune.ac.in

Authors

Sandhya Bhatia − *National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560 065, India; Indian Institute of Science Education and Research, Pune 411 008, India;* orcid.org/0000-0002-0926-8287

Guruswamy Krishnamoorthy − *Department of Biotechnology, Anna University, Chennai 600 025, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.0c11097

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank 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 Science and Technology, Government of India.

REFERENCES

(1) Udgaonkar, J. B. Multiple Routes and Structural Heterogeneity in Protein Folding. *Annu. Rev. Biophys.* **2008**, *37* (1), 489−510.

(2) Rollins, G. C.; Dill, K. A. General Mechanism of Two-State Protein Folding Kinetics. *J. Am. Chem. Soc.* **2014**, *136* (32), 11420−11427.

(3) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science (Washington, DC, U. S.)* **2012**, *338* (6110), 1042−1046.

(4) Malhotra, P.; Udgaonkar, J. B. How Cooperative Are Protein Folding and Unfolding Transitions? *Protein Sci.* **2016**, *25* (11), 1924−1941.

(5) Finkelstein, A. V. 50+ Years of Protein Folding. *Biochemistry (Moscow)* **2018**, *83*, S3−S18.

(6) Baldwin, R. L. The Nature of Protein Folding Pathways: The Classical versus the New View. *J. Biomol. NMR* **1995**, *5*, 103−109.

(7) Englander, S. W.; Mayne, L. The Case for Defined Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (31), 8253−8258.

(8) Eaton, W. A.; Wolynes, P. G. Theory, Simulations, and Experiments Show That Proteins Fold by Multiple Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (46), E9759−E9760.

(9) Baldwin, R. L. Clash between Energy Landscape Theory and Foldon-Dependent Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (32), 8442−8443.

(10) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein Folding Intermediates: Native-State Hydrogen Exchange. *Science (Washington, DC, U. S.)* **1995**, *269* (5221), 192−197.

(11) Baldwin, R. L.; Rose, G. D. Is Protein Folding Hierarchic? I. Local Structure and Peptide Folding. *Trends Biochem. Sci.* **1999**, *24* (1), 26−33.

(12) Adhikari, A. N.; Freed, K. F.; Sosnick, T. R. De Novo Prediction of Protein Folding Pathways and Structure Using the Principle of Sequential Stabilization. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (43), 17442−17447.

(13) Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (45), 15873−15880.

(14) Hu, W.; Kan, Z. Y.; Mayne, L.; Englander, S. W. Cytochrome c Folds through Foldon-Dependent Native-like Intermediates in an Ordered Pathway. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (14), 3809−3814.

(15) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21* (3), 167−195.

(16) Wolynes, P.; Onuchic, J.; Thirumalai, D. Navigating the Folding Routes. *Science (Washington, DC, U. S.)* **1995**, *267* (5204), 1619−1620.

(17) Weissman, J. S. All Roads Lead to Rome? The Multiple Pathways of Protein Folding. *Chem. Biol.* **1995**, *2* (5), 255−260.

(18) Ozkan, S. B.; Wu, G. A.; Chodera, J. D.; Dill, K. A. Protein Folding by Zipping and Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (29), 11987−11992.

(19) Harrison, S. C.; Durbin, R. Is There a Single Pathway for the Folding of a Polypeptide Chain? *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82* (12), 4028−4030.

(20) Ellison, P. A.; Cavagnero, S. Role of Unfolded State Heterogeneity and En-Route Ruggedness in Protein Folding Kinetics. *Protein Sci.* **2006**, *15* (3), 564−582.
(21) Hartl, F. U.; Hayer-Hartl, M. Converging Concepts of Protein Folding in Vitro and in Vivo. *Nat. Struct. Mol. Biol.* **2009**, *16* (6), 574−581.
(22) Gruebele, M.; Dave, K.; Sukenik, S. Globular Protein Folding In Vitro and In Vivo. *Annu. Rev. Biophys.* **2016**, *45* (1), 233−251.
(23) Anunciado, D. B.; Nyugen, V. P.; Hurst, G. B.; Doktycz, M. J.; Urban, V.; Langan, P.; Mamontov, E.; Neill, H. O. In Vivo Protein Dynamics on the Nanometer Length Scale and Nanosecond Time Scale. *J. Phys. Chem. Lett.* **2017**, *8* (8), 1899−1904.
(24) Radford, S. E.; Dobson, C. M.; Evans, P. A. The Folding of Hen Lysozyme Involves Partially Structured Intermediates and Multiple Pathways. *Nature* **1992**, *358* (6384), 302−307.
(25) Shastry, M. C. R.; Udgaonkar, J. B. The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* **1995**, *247*, 1013−1027.
(26) Goldbeck, R. A.; Thomas, Y. G.; Chen, E.; Esquerra, R. M.; Kliger, D. S. Multiple Pathways on a Protein-Folding Energy Landscape: Kinetic Evidence. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (6), 2782−2787.
(27) Rami, B. R.; Udgaonkar, J. B. pH-Jump-Induced Folding and Unfolding Studies of Barstar: Evidence for Multiple Folding and Unfolding Pathways. *Biochemistry* **2001**, *40* (50), 15267−15279.
(28) Krantz, B. A.; Dothager, R. S.; Sosnick, T. R. Discerning the Structure and Energy of Multiple Transition States in Protein Folding Using ψ-Analysis. *J. Mol. Biol.* **2004**, *337* (2), 463−475.
(29) Patra, A. K.; Udgaonkar, J. B. Characterization of the Folding and Unfolding Reactions of Single-Chain Monellin: Evidence for Multiple Intermediates and Competing Pathways. *Biochemistry* **2007**, *46* (42), 11727−11743.
(30) Jha, S. K.; Dasgupta, A.; Malhotra, P.; Udgaonkar, J. B. Identification of Multiple Folding Pathways of Monellin Using Pulsed Thiol Labeling and Mass Spectrometry. *Biochemistry* **2011**, *50* (15), 3062−3074.
(31) Charlier, C.; Reid Alderson, T.; Courtney, J. M.; Ying, J.; Anfinrud, P.; Bax, A. Study of Protein Folding under Native Conditions by Rapidly Switching the Hydrostatic Pressure inside an NMR Sample Cell. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (18), E4169−E4178.
(32) Viguera, A. R.; Blanco, F. J.; Serrano, L. The Order of Secondary Structure Elements Does Not Determine the Structure of a Protein but Does Affect Its Folding Kinetics. *J. Mol. Biol.* **1995**, *247* (4), 670−681.
(33) Bhuyan, A. K.; Udgaonkar, J. B. Observation of multistate kinetics during the slow folding and unfolding of barstar. *Biochemistry* **1999**, *38*, 9158−9168.
(34) Lindberg, M.; Tångrot, J.; Oliveberg, M. Complete Change of the Protein Folding Transition State upon Circular Permutation. *Nat. Struct. Biol.* **2002**, *9* (11), 818−822.
(35) Wright, C. F.; Lindorff-Larsen, K.; Randles, L. G.; Clarke, J. Parallel Protein-Unfolding Pathways Revealed and Mapped. *Nat. Struct. Mol. Biol.* **2003**, *10* (8), 658−662.
(36) Zaidi, F. N.; Nath, U.; Udgaonkar, J. B. Multiple Intermediates and Transition States during Protein Unfolding. *Nat. Struct. Biol.* **1997**, *4* (12), 1016−1024.
(37) Aghera, N.; Udgaonkar, J. B. Kinetic Studies of the Folding of Heterodimeric Monellin: Evidence for Switching between Alternative Parallel Pathways. *J. Mol. Biol.* **2012**, *420* (3), 235−250.
(38) Moulick, R.; Goluguri, R. R.; Udgaonkar, J. B. Ruggedness in the Free Energy Landscape Dictates Misfolding of the Prion Protein. *J. Mol. Biol.* **2019**, *431* (4), 807−824.
(39) Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B. Protein Folding and Unfolding on a Complex Energy Landscape. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (6), 2527−2532.
(40) Olofsson, M.; Hansson, S.; Hedberg, L.; Logan, D. T.; Oliveberg, M. Folding of S6 Structures with Divergent Amino Acid Composition: Pathway Flexibility Within Partly Overlapping Foldons. *J. Mol. Biol.* **2007**, *365* (1), 237−248.
(41) Connell, K. B.; Miller, E. J.; Marqusee, S. The Folding Trajectory of RNase H Is Dominated by Its Topology and Not Local Stability: A Protein Engineering Study of Variants That Fold via Two-State and Three-State Mechanisms. *J. Mol. Biol.* **2009**, *391* (2), 450−460.
(42) Malhotra, P.; Udgaonkar, J. B. Secondary Structural Change Can Occur Diffusely and Not Modularly during Protein Folding and Unfolding Reactions. *J. Am. Chem. Soc.* **2016**, *138* (18), 5866−5878.
(43) Jethva, P. N.; Udgaonkar, J. B. Modulation of the Extent of Cooperative Structural Change during Protein Folding by Chemical Denaturant. *J. Phys. Chem. B* **2017**, *121* (35), 8263−8275.
(44) Pradeep, L.; Udgaonkar, J. B. Differential Salt-Induced Stabilization of Structure in the Initial Folding Intermediate Ensemble of Barstar. *J. Mol. Biol.* **2002**, *324* (2), 331−347.
(45) Sridevi, K.; Lakshmikanth, G. S.; Krishnamoorthy, G.; Udgaonkar, J. B. Increasing Stability Reduces Conformational Heterogeneity in a Protein Folding Intermediate Ensemble. *J. Mol. Biol.* **2004**, *337* (3), 699−711.
(46) Pirchi, M.; Ziv, G.; Riven, I.; Cohen, S. S.; Zohar, N.; Barak, Y.; Haran, G. Single-Molecule Fluorescence Spectroscopy Maps the Folding Landscape of a Large Protein. *Nat. Commun.* **2011**, *2* (1), 493−497.
(47) Jagannathan, B.; Elms, P. J.; Bustamante, C.; Marqusee, S. Direct Observation of a Force-Induced Switch in the Anisotropic Mechanical Unfolding Pathway of a Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (44), 17820−17825.
(48) Yu, H.; Liu, X.; Neupane, K.; Gupta, A. N.; Brigley, A. M.; Solanki, A.; Sosova, I.; Woodside, M. T. Direct Observation of Multiple Misfolding Pathways in a Single Prion Protein Molecule. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (14), 5283−5288.
(49) Kotamarthi, H. C.; Sharma, R.; Narayan, S.; Ray, S.; Ainavarapu, S. R. K. Multiple Unfolding Pathways of Leucine Binding Protein (LBP) Probed by Single-Molecule Force Spectroscopy (SMFS). *J. Am. Chem. Soc.* **2013**, *135* (39), 14768−14774.
(50) Guinn, E. J.; Jagannathan, B.; Marqusee, S. Single-Molecule Chemo-Mechanical Unfolding Reveals Multiple Transition State Barriers in a Small Single-Domain Protein. *Nat. Commun.* **2015**, *6*, 1−8.
(51) Yoo, J.; Louis, J. M.; Chung, H. S. Diverse Folding Pathways of HIV-1 Protease Monomer on a Rugged Energy Landscape. *Biophys. J.* **2019**, *117* (8), 1456−1466.
(52) Ishay Ben, E.; Hazan, G.; Rahamim, G.; Amir, D.; Haas, E.; Ishay Ben, E.; Hazan, G.; Rahamim, G.; Amir, D.; Haas, E. 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.* **2012**, *83*, 084301.
(53) Bhatia, S.; Krishnamoorthy, G.; Dhar, D.; Udgaonkar, J. B. Observation of Continuous Contraction and a Metastable Misfolded State during the Collapse and Folding of a Small Protein. *J. Mol. Biol.* **2019**, *431* (19), 3814−3826.
(54) Halloran, K. T.; Wang, Y.; Arora, K.; Chakravarthy, S.; Irving, T. C.; Bilsel, O.; Brooks, C. L.; Matthews, C. R. Frustration and Folding of a TIM Barrel Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (33), 201900880.
(55) Lakshmikanth, G. S.; Sridevi, K.; Krishnamoorthy, G.; Udgaonkar, J. B. Structure Is Lost Incrementally during the Unfolding of Barstar. *Nat. Struct. Biol.* **2001**, *8* (9), 799−804.
(56) Kishore, M.; Krishnamoorthy, G.; Udgaonkar, J. B. Critical evaluation of the two-state model describing the equilibrium unfolding of the PI3K SH3 domain by time-resolved fluorescence resonance energy transfer. *Biochemistry* **2013**, *52*, 9482−9496.
(57) Ratner, V.; Amir, D.; Kahana, E.; Haas, E. Fast Collapse but Slow Formation of Secondary Structure Elements in the Refolding Transition of E. Coli Adenylate Kinase. *J. Mol. Biol.* **2005**, *352* (3), 683−699.

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457

(58) Jha, S. K.; Dhar, D.; Krishnamoorthy, G.; Udgaonkar, J. B. Continuous Dissolution of Structure during the Unfolding of a Small Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (27), 11113−11118.
(59) Navon, A.; Ittah, V.; Landsman, P.; Scheraga, H. A.; Haas, E. Distributions of Intramolecular Distances in the Reduced and Denatured States of Bovine Pancreatic Ribonuclease A. Folding Initiation Structures in the C-Terminal Portions of the Reduced Protein. *Biochemistry* **2001**, *40* (1), 105−118.
(60) Bhatia, S.; Krishnamoorthy, G.; Udgaonkar, J. B. Site-Specific Time-Resolved FRET Reveals Local Variations in the Unfolding Mechanism in an Apparently Two-State Protein Unfolding Transition. *Phys. Chem. Chem. Phys.* **2018**, *20* (5), 3216−3232.
(61) Goluguri, R. R.; Udgaonkar, J. B. Rise of the Helix from a Collapsed Globule during the Folding of Monellin. *Biochemistry* **2015**, *54* (34), 5356−5365.
(62) Goluguri, R. R.; Udgaonkar, J. B. Microsecond Rearrangements of Hydrophobic Clusters in an Initially Collapsed Globule Prime Structure Formation during the Folding of a Small Protein. *J. Mol. Biol.* **2016**, *428* (15), 3102−3117.
(63) Creed, D. Photochemistry and photophysics of near UV absorbing amino acids. I. Tryptophan and its simple derivatives. *Photochem. Photobiol.* **1984**, *39*, 537−83.
(64) Phillips, L. A.; Webb, S. P.; Martinez, S. J.; Phleming, G. R.; Levy, D. H. Time resolved spectroscopy of tryptophan conformers in a supersonic jet. *J. Am. Chem. Soc.* **1988**, *110*, 1352−5.
(65) Gutfreund, H. *Kinetics for the Life Sciences: Receptors, Transmitters and Catalysts*; Cambridge University Press: Cambridge, 1995.
(66) Sinha, K. K.; Udgaonkar, J. B. Dissecting the non-specific and specific components of the initial folding reaction of barstar by multi-site FRET measurements. *J. Mol. Biol.* **2007**, *370*, 385−405.
(67) Bedard, S.; Krishna, M. M. G. G.; Mayne, L.; Englander, S. W. Protein Folding: Independent Unrelated Pathways or Predetermined Pathway with Optional Errors. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (20), 7182−7187.
(68) Llinas, M.; Gillespie, B.; Dahlquist, F. W.; Marqusee, S. The Energetics of T4 Lysozyme Reveal a Hierarchy of Conformations. *Nat. Struct. Biol.* **1999**, *6* (11), 1072−1078.
(69) Plaxco, K. W.; Simons, K. T.; Baker, D. Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol.* **1998**, *277* (4), 985−994.
(70) Go, N. Theoretical Studies of Protein Folding. *Annu. Rev. Biophys. Bioeng.* **1983**, *12* (1), 183−210.
(71) Go, N. The Consistency Principle in Protein Structure and Pathways of Folding. *Adv. Biophys.* **1984**, *18* (C), 149−164.
(72) Mateos, B.; Conrad-Billroth, C.; Schiavina, M.; Beier, A.; Kontaxis, G.; Konrat, R.; Felli, I. C.; Pierattelli, R. The Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder Promoters to Compaction Facilitators. *J. Mol. Biol.* **2020**, *432* (9), 3093−3111.

https://dx.doi.org/10.1021/jacs.0c11097
J. Am. Chem. Soc. 2021, 143, 1447−1457