doi:10.1016/j.jmb.2012.04.019 J. Mol. Biol. (2012) 420, 235–250

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Kinetic Studies of the Folding of Heterodimeric Monellin: Evidence for Switching between Alternative Parallel Pathways

Nilesh Aghera and Jayant B. Udgaonkar*

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India

Received 23 September 2011; received in revised form 14 April 2012; accepted 18 April 2012
Available online 24 April 2012

Edited by C. R. Matthews

Keywords:
monellin;
heterodimeric protein;
folding kinetics;
parallel pathways

Determining whether or not a protein uses multiple pathways to fold is an important goal in protein folding studies. When multiple pathways are present, defined by transition states that differ in their compactness and structure but not significantly in energy, they may manifest themselves by causing the dependence on denaturant concentration of the logarithm of the observed rate constant of folding to have an upward curvature. In this study, the folding mechanism of heterodimeric monellin [double-chain monellin (dcMN)] has been studied over a range of protein and guanidine hydrochloride (GdnHCl) concentrations, using the intrinsic tryptophan fluorescence of the protein as the probe for the folding reaction. Refolding is shown to occur in multiple kinetic phases. In the first stage of refolding, which is silent to any change in intrinsic fluorescence, the two chains of monellin bind to one another to form an encounter complex. Interrupted folding experiments show that the initial encounter complex folds to native dcMN *via* two folding routes. A productive folding intermediate population is identified on one route but not on both of these routes. Two intermediate subpopulations appear to form in a fast kinetic phase, and native dcMN forms in a slow kinetic phase. The chevron arms for both the fast and slow phases of refolding are shown to have upward curvatures, suggesting that at least two pathways each defined by a different intermediate are operational during these kinetic phases of structure formation. Refolding switches from one pathway to the other as the GdnHCl concentration is increased.

Introduction

It has been argued that evolution must have favored multiple pathways, rather than only a single pathway, for an amino acid sequence to utilize in order to fold to a unique protein structure.[1] The availability of multiple folding pathways adds robustness to the protein folding reaction, and the utilization of more than one pathway would speed protein folding.[2,3] Multiple pathways have indeed been observed to operate during the folding and unfolding of several proteins.[4–12] Much is yet to be learnt about how competing folding pathways differ in the sequence of structural events defined by each pathway, as well as about the nature of the free energy barriers on a folding pathway.[13,14]

Multiple folding pathways are associated with structural heterogeneity in the unfolded state[15–18] and intermediates.[19,20] They arise when interconversion between subpopulations in these states is slow compared to folding. The important question really is: how and when during folding is structural heterogeneity reduced?[3] Protein folding pathways

*Corresponding author. E-mail address: jayant@ncbs.res.in.

Abbreviations used: dcMN, double-chain monellin; scMN, single-chain monellin; TS, transition state; GdnHCl, guanidine hydrochloride.

have been likened to 'funnels rather than tunnels' in conformational space,[21] and funnel pictures of protein folding are useful for depicting how conformational heterogeneity can reduce during folding reactions.[22,23] Experimental identification of multiple pathways is difficult because of the ensemble-averaging nature of the probes typically used to follow folding and unfolding reactions,[3,24] and because of the difficulty in identifying folding conditions in which structural heterogeneity becomes observable. A definitive but unusual signature of multiple pathways, upward curvature in the dependence on denaturant concentration of the logarithm of the folding/unfolding rate, has been observed in kinetic studies of the unfolding of only one protein.[11] Not surprisingly, multiple folding pathways have been detected experimentally for only a few proteins.

The folding of heterodimeric proteins can be carried out in the absence of any denaturant, providing an opportunity to study whether multiple pathways are more easily discernible in the most native-like conditions possible. The majority of all proteins are dimeric or multimeric,[25,26] but much of what is known about how a polypeptide chain self-assembles into a folded structure comes from studies with monomeric proteins.[3,13,27,28] The study of the folding mechanisms of multimeric proteins has, however, gained momentum[9,29–32] and is expected to lead to a better understanding of protein–protein interactions. When the individual subunits are unstructured prior to association, dimeric proteins provide an opportunity to study how folding is coupled to binding, which is important for understanding the functioning of intrinsically disordered proteins,[33] as well as the assembly of large protein assemblies.[34] In this context, it becomes important to study the mechanism of folding of heterodimeric proteins, especially those proteins for which both double-chain and single-chain variants possessing the same fold are available. The larger entropic barriers to folding for the double-chain variant than for the single-chain variant might lead to transition state (TS) structures, but not TS energies, on multiple pathways being sufficiently different that the multiple pathways manifest themselves in kinetic studies.

One such heterodimeric protein is the sweet protein monellin. Naturally occurring monellin [double-chain monellin (dcMN)] is a heterodimer of two subunits, chain A and chain B, which associate through interchain noncovalent interactions[35–37] (Fig. 1). A single-chain variant of monellin (scMN) was engineered at the genetic level, to create a protein in which the C-terminus of chain B was connected to the N-terminus of chain A using a Gly-Phe dipeptide linker.[38] The structures of dcMN and scMN are almost identical.[38,39] The study of the folding of scMN has yielded much useful insights about chain collapse[40,41] and the utilization of multiple pathways during folding.[7,10,42] The study of the unfolding of scMN has shown that unfolding commences through the formation of a dry molten globule[43] and proceeds gradually and not in an all-or-none manner.[44] Until recently, there were only a few quantitative studies of the unfolding of dcMN at equilibrium,[37,45] but the recent development of a good expression system for dcMN[46] made it possible to make a detailed study of its unfolding at equilibrium in relation to that of scMN.[39] This has set the stage for quantitative studies of its folding mechanism.

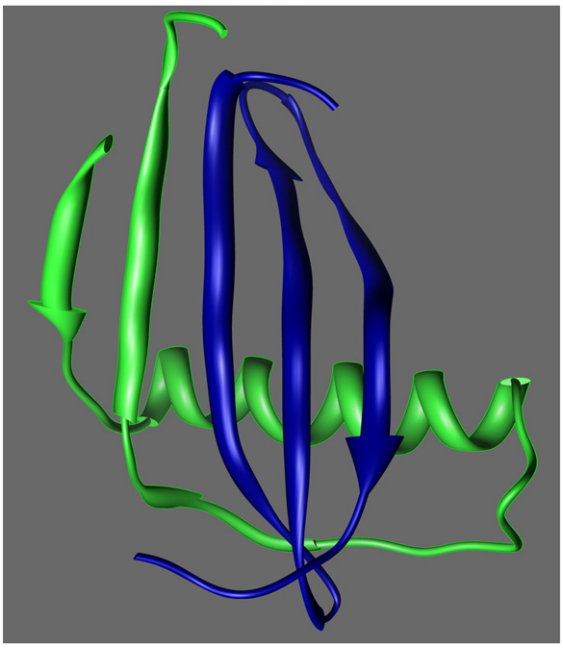


Fig. 1. Structure of dcMN. Chain A, consisting of 45 amino acid residues, is shown in blue, and chain B, consisting of 50 amino acid residues, is shown in green. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by National Institutes of Health grant P41 RR001081).

In this study, the refolding kinetics of dcMN have been studied over a range of protein concentrations, as well as over a range of low guanidine hydrochloride (GdnHCl) concentrations, using the intrinsic tryptophan fluorescence of the protein as a probe. The kinetics show three observable phases, each dependent on protein concentration as a consequence of folding reactions being coupled to the initial chain A–chain B binding reaction that leads to the formation of an encounter complex. Very unusually, plots of log (observed rate) *versus* GdnHCl concentration for both the fast and slow phases of refolding show upward curvatures, indicating that multiple pathways are operational during both folding phases. Interrupted refolding experiments indicate that native protein can form not only directly from the encounter complex

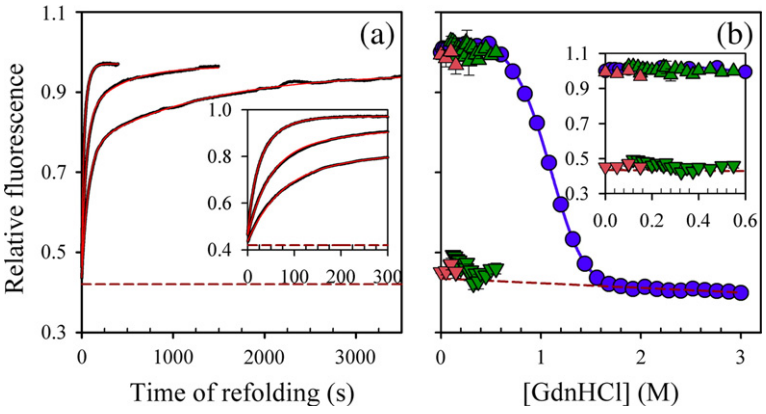


Fig. 2. Effect of GdnHCl concentration on the refolding kinetics of dcMN, monitored by the change in intrinsic tryptophan fluorescence at 340 nm at pH 7, 25 °C. (a) Representative kinetic refolding traces (continuous black line) when unfolded protein in 2.5 M GdnHCl was refolded by diluting the GdnHCl to 0.1, 0.2 and 0.3 M (top to bottom). The continuous red lines are least-squares fits of the data to a two-exponential equation. The inset shows the initial parts of the kinetic traces. (b) Kinetic *versus* equilibrium amplitudes of refolding. Kinetic experiments were either folding (mixing chain A with chain B in the presence of different concentrations of GdnHCl) or refolding (diluting protein unfolded in 2.5 M GdnHCl to different final GdnHCl concentrations) experiments. ●, equilibrium unfolding curve; ▼, $t=0$ points of the refolding kinetic traces; ▼, $t=0$ points of folding kinetic traces; ▲, $t=\infty$ points of the refolding kinetic traces; ▲, $t=\infty$ points of folding kinetic traces. The blue continuous line through the equilibrium unfolding data is a nonlinear, least-squares fit to a two-state N↔U model.[39] The inset shows the data for the lowest GdnHCl concentrations. The error bars, wherever shown, represent the spreads of measurements from at least two separate experiments. In both panels, the broken line is an extrapolation of the unfolded protein baseline from equilibrium unfolding studies. In all the experiments, the final concentration of dcMN was 10 μM.

but also indirectly *via* folding intermediates on two separate pathways. A minimal kinetic mechanism is proposed to account for the kinetic data, and the mechanism has been validated by global fitting of the kinetic traces obtained at all GdnHCl concentrations.

Results

In this study, the reaction leading to the formation of native protein was commenced in two different ways. In the refolding by denaturant dilution experiment, protein that had been unfolded in 2.5 M GdnHCl was refolded by dilution of the denaturant. In the folding by chain complementation experiment, free chains in native buffer were allowed to undergo chain complementation in different concentrations of denaturant. The folding experiment allows the formation of native structure to be studied in the absence of denaturant and at very low denaturant concentrations unachievable in denaturant dilution experiments.

Refolding occurs in two kinetic phases whose chevron arms have upward curvatures

Refolding by denaturant dilution of 10 μM dcMN occurs in two kinetic phases, one fast and one slow (Fig. 2a), at all GdnHCl concentrations studied. The total amplitude of fluorescence change for both kinetic phases is the same as the amplitude observed in equilibrium unfolding/refolding studies (Fig. 2b), indicating that the observed fast and slow kinetic phases account for the entire refolding reaction.

Quite remarkably, for both kinetic phases, the dependence on GdnHCl concentration of the logarithm of the observed rate constant shows an upward curvature (Fig. 3b and c). The upward curvature is more pronounced for the slow phase, with a kink being apparent in the refolding chevron arm at about 0.2 M GdnHCl. It should be noted that refolding was initiated by manual mixing for GdnHCl concentrations in the range of 0.25 – 0.55 M and by stopped-flow mixing for GdnHCl concentrations from 0.1 M to 0.45 M and that the rate constants obtained from both mixing methods are identical in the overlapping range of denaturant concentration, thereby ruling out the possibility that the upward curvature is an experimental artifact. For both the fast and slow kinetic phases, the dependences of the rate constants on GdnHCl concentrations are seen to fit well to a simple equation based on refolding occurring on two parallel pathways (Fig. 3b and c). The relative amplitude of the fast phase increases at the expense of the slow phase as the GdnHCl concentration is raised from 0.1 to about 0.2 M, after which the relative amplitude of the slow phase increases at the expense of the fast phase (Fig. 3e and f).

Since the two chains, A and B, of dcMN are mostly unstructured in isolation in the absence of denaturant,[39,47] it was possible to initiate the folding of dcMN by mixing the two chains together in an equimolar ratio, at different very low (0 to 0.15 M) GdnHCl concentrations. In such folding by chain complementation studies carried out at the same final protein concentration (10 μM), as were the refolding by denaturant dilution studies described above, three kinetic phases are observed, very fast, fast, and slow. As with the refolding by denaturant dilution studies, no burst phase is observable in the dead time of stopped-flow mixing. The rate

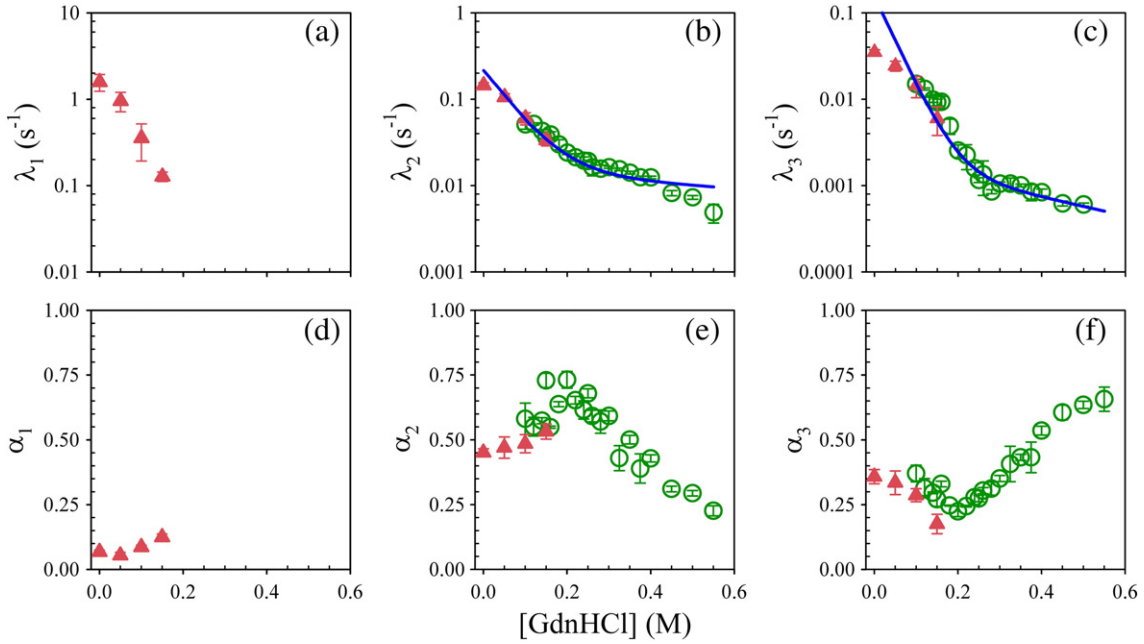


Fig. 3. Dependence of the observed rate constants and relative amplitudes on GdnHCl concentration. Denaturant-dependent folding/refolding kinetic studies were performed at 10 μM dcMN concentration and were monitored by the intrinsic tryptophan fluorescence change at 340 nm. Folding rate constants and relative amplitudes (▲) were obtained from three-exponential fits of the kinetic traces; refolding rate constants and relative amplitudes (○) were obtained from two-exponential fits of the kinetic traces. (a–c) The observed rate constants for the very fast, fast, and slow phases of folding/refolding, respectively. (d–f) The relative amplitudes of the very fast, fast, and slow phases of folding/refolding, respectively. The blue continuous lines in (b) and (c) were drawn using Eq. (2). The blue line in (b) was drawn using the following values for the different parameters: $\lambda_{2a}=0.2\ s^{-1}$, $m_{2a}=15\ M^{-1}$, $\lambda_{2b}=0.015\ s^{-1}$, and $m_{2b}=0.081\ M^{-1}$. The blue line in (c) was drawn using the following values for the different parameters: $\lambda_{3a}=0.15\ s^{-1}$, $m_{3a}=24\ M^{-1}$, $\lambda_{3b}=0.002\ s^{-1}$, and $m_{3b}=2.5\ M^{-1}$. In all panels, the error bars represent the spreads of measurements from at least two separate experiments.

constants and relative amplitudes of the fast and slow phases of folding by chain complementation merge with those of the corresponding fast and slow phase of refolding initiated by denaturant dilution, measured at higher GdnHCl concentrations (Fig. 3b, c, e, and f). The observed rate constant of the very fast phase becomes very similar in value to that of the fast phase at higher GdnHCl concentrations, but for consistency, all the kinetic data for refolding initiated by denaturant dilution were analyzed with a two-exponential equation, whereas all kinetic data for folding by chain complementation were analyzed with a three-exponential equation.

Kinetics of folding by chain complementation depends on protein concentration

To delineate the binding and structure acquisition events during the folding of dcMN, studies of folding by chain complementation were carried out at different initial protein concentrations from 2 to 30 μM of each chain. The binding step, a second-order reaction, is expected to be dependent on protein concentration. When folding is carried out in the absence of GdnHCl, the very fast, fast, and slow phases of folding are seen at all protein concentrations, and no burst phase is seen (Fig. 4). The rate constants observed for all three kinetic phases of folding are dependent on protein concentration, but the dependences are much less than expected: the rate constant observed for the very fast phase increases about 3-fold (Fig. 4b), and those for the fast (Fig. 4c) and slow (Fig. 4d) phases increase about 9-fold for a 15-fold increase in protein concentration. The relative amplitudes of the three phases do not change significantly over the range of protein concentrations studied (Fig. 4e).

Refolding kinetics depends on protein concentration

When refolding is initiated from protein unfolded in 2.5 M GdnHCl, either in 0.25 M GdnHCl (Fig. 5a) or in 0.4 M GdnHCl (Fig. 5b), the fast and slow phases are seen at all protein concentrations, and no burst phase or very fast phase is seen. At 0.25 M GdnHCl, where the dependence on GdnHCl concentration of the logarithm of the fast or slow rate

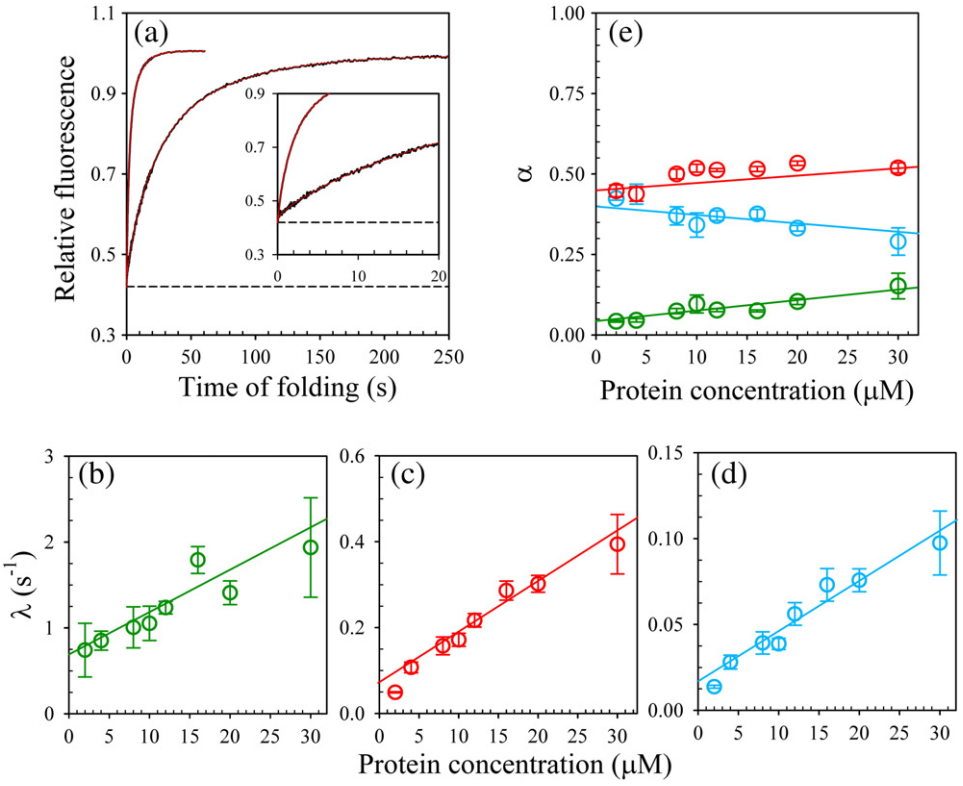


Fig. 4. Dependence of the folding kinetics of dcMN on protein concentration at pH 7, 25 °C. (a) Folding kinetic traces (black lines) were obtained by mixing equimolar concentrations of chains A and B in the absence of GdnHCl, yielding 2 μM (right trace) and 30 μM (left trace) dcMN. The inset shows the initial parts of the kinetic traces. The red lines are least-squares fits to a three-exponential equation. The broken line represents the fluorescence of unfolded dcMN, obtained from equilibrium unfolding studies. (b–d) The dependences of the rate constants observed for the very fast (○), fast (○), and slow (○) phases of folding, respectively, on protein concentration. (e) The dependences of the relative amplitudes of the three kinetic phases on protein concentration. In (b)–(e), the continuous lines have been drawn by inspection only, and the error bars represent the spread in values obtained from two different experiments.

constant shows its upward curvature (kink), only the rate constant for the fast phase shows a dependence on protein concentration, while that of the slow phase is observed to be independent of protein concentration (Fig. 5c and e). At 0.4 M GdnHCl, the rate constants observed for both the fast and the slow phases are weakly dependent on protein concentration (Fig. 5d and f). At both GdnHCl concentrations, the relative amplitude of the fast phase of refolding increases at the expense of that of the slow phase, with an increase in protein concentration, and at any one protein concentration, the relative amplitudes of both phases are different for refolding at the different GdnHCl concentrations (Fig. 5g and h).

Interrupted refolding experiments indicate the formation of an intermediate and two major folding routes

Refolding initiated by denaturant dilution from 2.5 to 0.25 M GdnHCl was interrupted at different times in the range 0 to 3000 s by a jump in GdnHCl concentration to 5 M to initiate unfolding. The resultant unfolding trace when extrapolated to $t=0$ was found not to originate at the fluorescence value predicted by the kinetic trace of refolding (Fig. 6a). This indicated that the fluorescence change during unfolding occurs in two kinetic phases: a burst phase that is not observable in the 10 s dead time of manual mixing and a slower phase occurring at a measurable rate of unfolding (0.02 ± 0.001 s^{-1}). Identical results were obtained when the interrupted refolding studies were carried out using stopped-flow mixing with a dead time of 100 ms.

The slower phase occurs at a rate corresponding to the rate of unfolding of fully native protein (N)[39] and, hence, reflects the unfolding of protein that has fully folded to N at the time of application of the unfolding jump. Its observed amplitude therefore indicates the amount of N present at the time refolding was interrupted by the unfolding jump. The relative amount of N increases in a single kinetic phase with a rate constant of 0.002 ± 0.0002 s^{-1} (Fig.

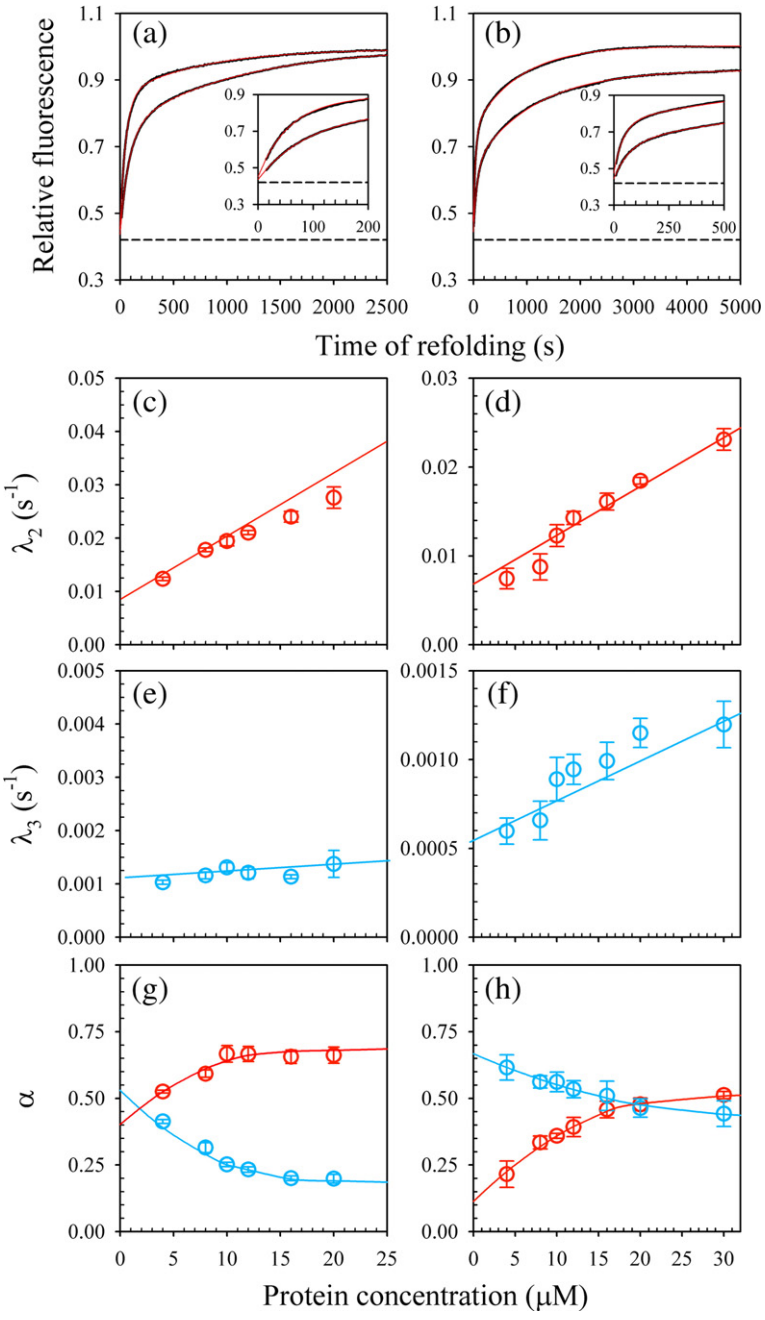


Fig. 5. Dependence of the refolding kinetics of dcMN on protein concentration at pH 7, 25 °C. Refolding kinetic traces (black lines) were obtained by diluting unfolded protein in 2.5 M GdnHCl to 0.25 M (a) and 0.4 M (b) GdnHCl. Refolding traces for 4 μM (right trace) and 20 μM (left trace) protein are shown in (a), and those for 4 μM (right trace) and 30 μM (left trace) protein are shown in (b). The insets show the initial parts of the kinetic traces. The red lines are least-squares fits to a two-exponential equation. The broken line represents the fluorescence of unfolded dcMN, obtained from equilibrium unfolding studies. (c) and (e) show the rate constants for the fast (o) and slow (o) phases of refolding in 0.25 M GdnHCl, while (g) shows the relative amplitudes of the two kinetic phases. (d) and (f) show the rate constants for the fast (o) and slow (o) phases of refolding in 0.4 M GdnHCl, while (h) shows the relative amplitudes of the two kinetic phases. In (c)–(h), the continuous lines have been drawn by inspection only, and the error bars represent the spread in values obtained from two different experiments.

6b). This rate constant is the same as the rate constant of the slow phase of refolding in 0.25 M GdnHCl. No lag phase is observed in the formation of N.

Since a burst-phase change in fluorescence is not observed during the unfolding of native dcMN,[39] the burst phase of unfolding observed for partially folded protein must represent the unfolding of at least one partially folded intermediate, I, populated transiently during the refolding reaction. The relative amplitude of the burst-phase change in fluorescence increases and then decreases in value, as the time at which refolding is interrupted by the unfolding jump is increased (Fig. 6c). Hence, the intermediate, which is so unstable that it unfolds fully within a 100-ms burst phase, increases in population with time of refolding up to 100 s and then decreases in population to nearly zero at 2000 s of refolding. The apparent rate of formation of the intermediate is $0.02 \pm 0.002\ s^{-1}$ and the apparent rate of its disappearance is $0.002 \pm 0.0002\ s^{-1}$. The observed rate constant for the formation of the intermediate is the same as the observed rate constant of the fast phase of refolding kinetics at

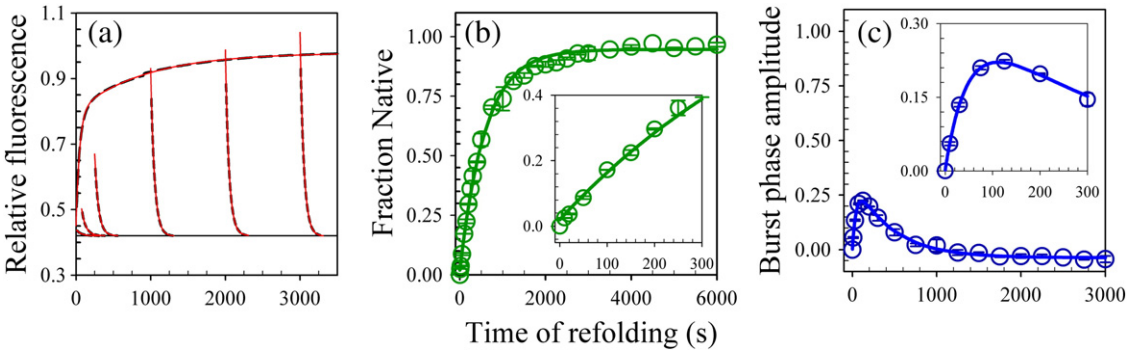


Fig. 6. Interrupted refolding experiments. (a) Kinetic traces of unfolding in 5 M GdnHCl, initiated at different times of refolding of 10 μM dcMN in 0.25 M GdnHCl at pH7 (short black broken lines). Refolding was interrupted at 10, 75, 250, 1000, 2000, and 3000 s (left to right). The red continuous lines through the unfolding traces represent least-squares fits to a single-exponential equation, extrapolated to the $t=0$ points of the kinetic unfolding traces. Each unfolding trace was found to decay with a rate of 0.02 ± 0.0017 s^{-1}. Also shown is the kinetic trace of refolding of 10 μM dcMN in 0.25 M GdnHCl at pH7 (long black broken line), through which is drawn a two-exponential least-squares fit (red continuous line). All the kinetic traces were normalized to the $t=\infty$ point of the unfolding traces. (b) Kinetics of formation of N. The amplitude of the single-exponential fit to the kinetic trace of unfolding initiated at each time of refolding, relative to the amplitude of the single-exponential fit to the kinetic trace of unfolding of fully folded protein, is plotted against the time of refolding. The inset shows the first 300 s of refolding. The continuous line through the data represents a least-squares fit to a single-exponential equation and yields an apparent rate constant of 0.002 ± 0.0001 s^{-1} for the formation of N. (c) Kinetics of formation of intermediate I. The difference between the fluorescence value of the refolding trace at a particular time point and the fluorescence value represented by the extrapolated $t=0$ point of the kinetic unfolding trace originating from the same time point of refolding, relative to the total fluorescence change observed during the unfolding of fully folded protein, is plotted against time of refolding. The inset shows data for the first 300 s. The continuous line through the data is a fit to a sequential three-state model [Eq. (1)]. The signal increases at a rate of the 0.02 ± 0.001 s^{-1} and then decreases at a rate of 0.002 ± 0.0001 s^{-1}. In (b) and (c), the error bars represent the spreads of measurements made in two separate experiments.

0.25 M GdnHCl. The observed rate constant for the disappearance of the intermediate is the same as that of the slow phase of refolding in 0.25 M GdnHCl, which is the same as the rate of formation of N (Fig. 6b).

Discussion

In this study of the folding mechanism of dcMN, folding by chain complementation experiments were carried out by mixing the isolated free chains A and B in 0 to 0.15 M GdnHCl, and refolding by denaturant dilution experiments were carried out by diluting dcMN that had been unfolded in 2.5 M GdnHCl to different low GdnHCl concentrations. In the former case, a very fast, a fast and a slow phase of folding are observed, while in the latter case, only the fast and slow phases are observed. The observation that the dependences on GdnHCl concentration of the rate constants, as well as of the relative amplitudes obtained for the folding experiments starting from the isolated chains in native conditions, merge with the corresponding dependences for the refolding experiments starting from protein unfolded in 2.5 M GdnHCl suggests that both folding and refolding are describable by the same folding mechanism. The observation of three kinetic phases for the folding by chain complementation experiments indicates that at least four distinct conformations, including the unfolded and native conformations, must be incorporated in the folding mechanism; hence, the folding mechanism must describe a chain A–chain B association step as well as at least two structure acquisition steps.

For some dimeric proteins, chain–chain association is the first step in folding,[29,31,32] presumably because the folding code is resident in both chains. For others, the chain–chain association step occurs after initial folding of the two chains;[9,30] this has the advantage of minimizing potential aggregation of the free chains. In some cases, it has not been possible to temporally dissect out the binding and folding events,[48] but it is very unlikely that the binding and folding events occur concurrently. In the case of dcMN, it is known that chains A and B are essentially unstructured in isolation in folding conditions[39,47] and that fluorescence of the chains does not change when transferred from folding to unfolding conditions (data not shown); hence, the first step in folding/refolding must be chain A–chain B association to form an encounter complex C.

Such an inference is supported by the observation that, although the apparent rates of the three phases of folding/refolding depend on protein concentration, except that of the slow phase of refolding in

0.25 M GdnHCl (Figs. 4 and 5), the dependences are weak. In particular, the very fast phase does not increase fourfold when the chain/protein concentration is increased twofold,[32,49] as would be expected if the very fast phase represented binding of chain A to chain B. Furthermore, a plot of the observed rate constant *versus* protein concentration does not extrapolate to the origin.[50] It appears therefore that binding precedes the very fast phase, and because no burst-phase change is observed during folding/refolding, the binding event, $A + B \leftrightarrow C$, must be silent to fluorescence change. The binding of chain A to chain B to form native dcMN is tight, with an equilibrium constant for dissociation of 40 nM;[39] hence, it is likely that the initial binding event is fast compared to the very fast phase of folding. Consequently, a pre-equilibrium would be established between A, B, and C before subsequent structure acquisition occurs, as has been reported for the association of SRP and its receptor SR.[32]

It should be noted that once chains A and B have associated to form C, the subsequent structure acquisition reactions are first-order reactions and are not expected to be dependent on protein concentration. However, the structure acquisition steps are kinetically coupled to the initial binding step, and hence, the observed rate constants for the very fast, fast, and slow phases will show weak dependences on protein concentration, depending on the values for the individual microscopic rate constants for folding as well as for binding (Figs. 4 and 5). The observation of such kinetic coupling suggests that at the concentration at which the folding/refolding experiments were carried out (10 μM), a substantial fraction of the protein must be present as free chains A and B, upon establishment of the initial pre-equilibrium between chains A and B and the encounter complex C. It would therefore appear that the equilibrium dissociation constant defining this pre-equilibrium is near about 10 μM.

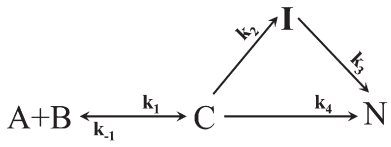
The folding of complex C to N proceeds through intermediate(s) I

The observation that the apparent rate of disappearance of I is the same as the rate of appearance of N (Fig. 6) indicates that I is a productive intermediate populated on the route from C to N, and that it is not a dead-end, off-pathway intermediate. The observation that the slow phase of refolding in 0.25 M GdnHCl does not depend on protein concentration (Fig. 5e) is also consistent with I not being an off-pathway, dead-end intermediate arising from C, because in that case, the slow phase (C-to-N conversion) would be expected to show a dependence on protein concentration. The observation that the apparent rate of the fast phase of refolding, as measured in direct refolding experiments, is the same as the apparent rate of formation of the intermediate I observed in interrupted refolding experiments (Fig. 6) suggests that the fast phase of refolding leads to the formation of I from C. The observation that the rate of the slow phase of refolding is the same as the rate of disappearance of I as well as the rate of formation of N suggests that the slow phase of refolding leads to the formation of N from I. These observations are consistent with an $A + B \leftrightarrow C \leftrightarrow I \leftrightarrow N$ scheme for folding/refolding, in which the fast phase of fluorescence change occurs during the $C \leftrightarrow I$ step and the slow phase occurs during the $I \leftrightarrow N$ step.

The observation that the relative amplitude of the fast phase decreases at the expense of the slow phase with an increase in GdnHCl concentration from 0.2 to 0.5 M (Fig. 3) can be accounted for by this scheme: the stability of I is expected to decrease with increasing GdnHCl concentration; hence, the extent to which it accumulates decreases with GdnHCl concentration. The surprising observation is that the relative amplitude of the fast phase decreases when the concentration of GdnHCl is reduced below about 0.2 M. A possible explanation is that the stability of I increases as the GdnHCl concentration is raised from 0 to 0.2 M, through an ionic strength effect wherein destabilizing electrostatic interactions are screened by the added salt. It is known that the folding kinetics of monellin is sensitive to the ionic strength and that folding slows down in the presence of salt,[37] supporting the possibility that I may be stabilized by an ionic strength effect exerted by low concentrations (< 0.2 M) of GdnHCl. Although native dcMN is not stabilized by low subdenaturing concentrations of GdnHCl (unpublished results), unlike other proteins,[51] it is possible that I is. Another possible explanation is that I is a structurally heterogeneous intermediate ensemble, composed of subpopulations of molecules with different fluorescence properties and that when the GdnHCl concentration is lowered below 0.2 M, a less-fluorescent subpopulation of the I ensemble gets preferentially stabilized. This would result in the relative amplitude of the fast phase of fluorescence change decreasing as the GdnHCl concentration present during refolding is lowered below 0.2 M GdnHCl. Previous studies have shown that denaturant concentration can modulate the structural heterogeneity of a folding intermediate[19] and that the structure manifested by an intermediate ensemble can be tuned by the addition of co-solutes such as salts and osmolytes.[52,53]

A direct C↔N route competes with the C↔I↔N route

I is maximally populated at 100 s of refolding, and a 100 s lag in the formation of N should have been observed if the formation of N occurs only *via*



Scheme 1.

the A + B ↔ C ↔ I ↔ N route. The interrupted refolding experiments that directly assay for N during refolding show that no such lag is present (Fig. 6b). Hence, it appears that C can also directly fold to N on a route, C ↔ N, competing with the C ↔ I ↔ N route (Scheme 1). The formation of N must occur at the same rate on both routes given that its kinetics is well described by a single-exponential equation (Fig. 6b). It should be noted that this observation only means that the free energy barriers separating C from N on the direct pathway, and I from N on the indirect pathway, are similar in magnitude, but it does not necessarily mean that the TSs atop these barriers have identical structures. A similar observation has been made in refolding studies of other proteins.[54,55] At present, the reason for multiple folding routes arising from the encounter complex C is not understood. It is possible that C itself is a heterogeneous mixture of subpopulations of molecules in slow equilibrium with each other and that the different folding routes arise from these different subpopulations.

Hence, the results discussed so far, from the direct folding/refolding experiments at different protein concentrations as well as from the interrupted refolding experiments, are well accounted for by the minimal kinetic mechanism depicted in Scheme 1. Such a triangular scheme in which the folding of an unstructured molecule (C) to N can occur either directly or indirectly *via* an intermediate population I is also the minimal scheme that describes the folding of lysozyme.[5] The observation here that the apparent rate of formation of N during refolding in 0.25 M GdnHCl is the same on both folding routes brings into focus what the role of I in folding might be, especially given that the rate of formation of I is 10-fold faster than that of N (see above), which means that the folding flux occurs predominantly *via* I. It is likely that I plays a role in reducing the entropy of activation, by partitioning the unfavorable entropy change accompanying folding into two steps.

Scheme 1 does not, however, explain the upward curvature seen in the chevron arms for both the fast and slow observed rate constants (Fig. 3b and c).

Curvature in a chevron arm

The free energy change for folding is known to be linearly dependent on denaturant concentration.[56–58] Hence, for the direct C ↔ N route in Scheme 1, the logarithm of the rate constant of folding is expected to have a linear dependence (*m* value) on denaturant concentration. Similarly, for the C ↔ I ↔ N route in Scheme 1, the logarithm of the individual rate constant for each step of refolding would have a linear dependence on denaturant concentration. Since structure formation most likely occurs in a progressive manner on any folding pathway, the TS for the first step is expected to be less compact and structured than the TS for the second step. This would lead to the dependence on denaturant concentration of the rate constant for the first fast step being less than that for the second step; consequently, the observed kinetic *m* value for refolding would be lower at lower denaturant concentration. Hence, a downward curvature would be observed either upon a change in the rate-limiting step from the second step to the first step or upon accumulation of the intermediate. Indeed, such downward curvatures have been observed for the refolding and unfolding reactions of several proteins that proceed through on-pathway intermediates.[12,59–62] Other possible reasons for downward curvatures include (1) a Hammond shift in the position of the TS along the reaction coordinate,[63] (2) transient aggregation of the folding protein,[64] and (3) experimental underestimation of very fast refolding rate constants.[65] It should be noted that if I were an off-pathway and dead-end intermediate, possibly because it possesses nonnative interactions, the folding chevron arm would have curved downwards because the observed refolding rate would have actually decreased with a decrease in GdnHCl concentration.[66–68] In this study, it is observed that the chevron arm for the slow rate constant shows a slight downward curvature at GdnHCl concentrations below 0.1 M. This is likely to be because of accumulation of an on-pathway I. The distinguishing feature of the chevrons for both the fast and slow rate constants is, however, the upward curvature that manifests itself at GdnHCl concentrations below 0.25 M.

Upwardly curved chevrons are indicative of multiple folding pathways

When competing folding pathways are present, the relative utilization of any particular pathway depends on the overall rate of refolding along that pathway, which in turn is dependent on the stability of the TS for that pathway. The presence of two competing pathways becomes discernible only when the TSs (TS1 and TS2) on the two pathways differ significantly in their structure and compactness. If TS1 is more compact than TS2, its stability will decrease more than that of TS2 for the same increase in GdnHCl concentration.

Consequently, TS2 will become more stable than TS1 at high GdnHCl concentrations, because it has

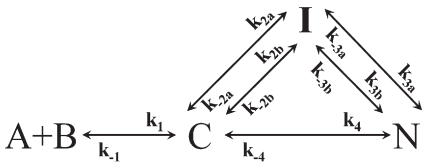
more solvent-accessible surface area for GdnHCl to interact with.[3,69] At low denaturant concentrations, refolding will occur along the pathway on which the TS (TS1) is compact and structured and hence relatively stable. At high denaturant concentrations, refolding will occur along the pathway on which the TS (TS2) is less compact and more unstructured than TS1. Because TS1 is more compact than TS2, the kinetic *m* value observed at low denaturant concentrations is larger than the *m* value observed at high denaturant concentrations. Hence, an upward curvature in the folding arm of the chevron will be observed when two competing pathways are present for which the TSs differ significantly in compactness. The upward curvatures observed for both the fast and slow refolding rate constant chevrons (Fig. 3) indicate therefore that the folding of dcMN switches from the TS1 pathway at GdnHCl concentrations below 0.2 M to the TS2 pathway at GdnHCl concentrations above 0.2 M.

The detection of multiple folding pathways in this manner for dcMN is very unusual and has not been reported before. On the other hand, multiple unfolding pathways for titin have indeed been detected on the basis of the unfolding arm of the chevron displaying upward curvature,[11] although no curvature was observed in the refolding arm of the chevron plot.[70] It should be noted that in this study of the folding of dcMN, the ability to initiate folding at very low GdnHCl concentration, by mixing together the free chains A and B, was instrumental in the detection of the parallel pathways. Switching between the alternative pathways is seen to occur at GdnHCl concentrations not usually utilized in the study of the refolding of monomeric proteins for which refolding is initiated by denaturant dilution.

Although an upwardly curved chevron must signify two or more competing pathways for refolding, the converse need not be true. When competing pathways are present, upward curvature in the chevron will be seen only when the TSs on the two pathways differ significantly in their compactness (see above), but not in their stabilities so that refolding switches from one pathway to the other with a change in denaturant concentration. If the TSs do not differ significantly in their compactness, an upward curvature in the chevron will not be seen. Indeed, for several proteins that have been shown to refold *via* multiple pathways, no upward curvature in the refolding arm of the chevron was observed.[5,8,42,61,71]

Refolding *via* intermediates occurs *via* two parallel pathways

The observation that the dependences on GdnHCl concentration of the logarithm of the apparent rate constant for the fast phase of refolding and for the slow phase of refolding show upward curvatures indicates that both steps on the intermediate route (C↔I↔N) occur *via* two pathways operating in parallel. I could be a single intermediate ensemble, in which case, there would be two pathways between C and I and two pathways between I and N (Scheme 2). Alternatively, I could consist of two subpopulations of intermediates, I_1 and I_2, which are formed simultaneously during the fast phase of folding on two competing pathways (Scheme 3).

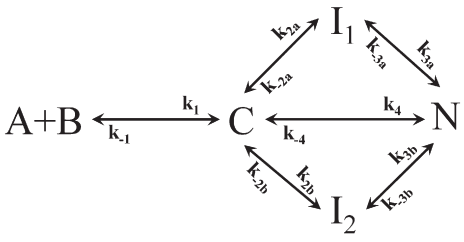


Scheme 2.

Distinguishing between Schemes 2 and 3, in both of which the two competing pathways starting from C would be defined by TS1 and TS2 (see above), is not easy. An important experimental observation that has to be explained is that the relative amplitude of the fast phase decreases for GdnHCl concentrations below 0.2 M GdnHCl, where switching from the TS1 pathway to the TS2 pathway also occurs (see above). Scheme 2, for which the equilibrium constant between C and I has to be the same for the two pathways, and for which I has the choice of utilizing either of two pathways to unfold, cannot explain this observation. Scheme 3 can explain this if I_1 is on the TS1 pathway, which is utilized more below 0.2 M GdnHCl than is the TS2/I_2 pathway (see above). Then, if the rate of formation of I_1 has a smaller dependence on GdnHCl concentration than does the rate of transformation of I_1 to N, I_1 would accumulate more in 0.2 M GdnHCl than it would at lower GdnHCl concentrations. Above 0.2 M GdnHCl, folding switches to the TS2/I_2 pathway (see above), and if on this pathway, the rate of formation of I_2 decreases less with an increase in GdnHCl concentration than does the rate of formation of N from I_2, then I_2 would accumulate less at higher GdnHCl concentrations. Hence, the relative amplitude of the fast phase would increase as the GdnHCl concentration is increased to 0.2 M and would decrease thereafter. Hence, Scheme 3 can qualitatively account for the experimental observation, but a more quantitative analysis is necessary. Nevertheless, and not surprisingly perhaps, it has been shown that the fast phase seen in the folding of scMN also leads to the formation of two intermediates of differing stabilities on two competing pathways.[7,42]

Kinetic model for the folding of dcMN

A global fitting program, written in MATLAB, was used to quantitatively analyze the dependence



Scheme 3.

of the refolding kinetics on GdnHCl concentration, according to Scheme 3. Details about the fitting program are provided in Materials and Methods. The fits in Fig. 7 show that Scheme 3 can account for all aspects of the data, not only the dependence of the kinetics on GdnHCl concentration but also the dependence on protein concentration. The fitted parameters are shown in Table 1. Importantly, the values obtained for the parameters are consistent with the qualitative reasoning that led to the proposal of Scheme 3, as described above. In contrast, global fitting to Scheme 2 was not satisfactory.

Global fitting of raw kinetic data to a model such as that in Scheme 3 is especially appropriate for multichain folding reactions where the protein concentration-dependent steps are not expected to follow exponential behavior (even though they might appear to do so) and yield accurate rate constants. It is stressed that the model in Scheme 3 is a minimal model and that the main purpose of global fitting of the kinetic data has been to show that the data are consistent with the model. While the model provides a kinetic description of multiple folding pathways, it does not provide structural explanation, for example, of why the apparent rate constants of folding from C to I_1 and to I_2 have such strong dependences on GdnHCl concentration. More work is needed to provide better kinetic and structural descriptions of the data.

Table 1. Kinetic parameters obtained by global fitting of all kinetic folding/refolding traces obtained at different GdnHCl concentrations using free chain concentrations of 10 μM

Folding				Unfolding			
Rate constant (s^{-1})		*m* value (kcal mol^{-1} M^{-1})		Rate constant (s^{-1})		*m* value (kcal mol^{-1} M^{-1})	
k_1	10^7 (M^{-1} s^{-1})	m_1	0	k_{-1}	653	m_{-1}	0.002
k_{2a}	2.21	m_{2a}	−11.34	k_{-2a}	0.00059	m_{-2a}	2.79
k_{2b}	0.098	m_{2b}	−1.57	k_{-2b}	0.0034	m_{-2b}	3.27
k_{3a}	0.24	m_{3a}	−25.58	k_{-3a}	0	m_{-3a}	0
k_{3b}	0.042	m_{3b}	−5.37	k_{-3b}	0	m_{-3b}	0
k_4	0.056	m_4	−6.87	k_{-4}	0	m_{-4}	0

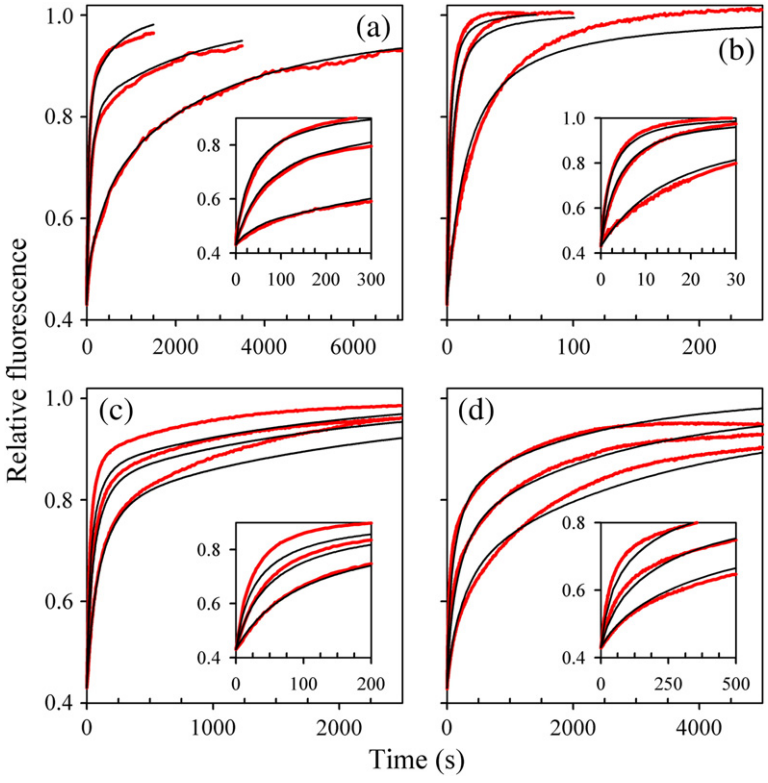


Fig. 7. Global fits of experimental kinetic traces to the mechanism in Scheme 2. A MATLAB program was used as described in Materials and Methods. In all panels, the experimental kinetic traces are shown as red continuous lines, and the fits for the same data from global fitting are shown as black continuous lines. (a) Representative experimental refolding kinetic traces and their fits obtained at 0.2, 0.3, and 0.5 M GdnHCl (left to right) at free chain concentrations of 10 μM. (b) Representative experimental folding kinetic traces and their fits obtained at 2, 10, and 30 μM chain concentrations (left to right) in the absence of GdnHCl. (c) Representative refolding kinetic traces and their fits obtained at 4, 10, and 20 μM free chain concentrations (left to right) in the presence of 0.25 M GdnHCl. (d) Representative refolding kinetic traces and their fits obtained at 4, 10, and 30 μM free chain concentrations (left to right) in the presence of 0.4 M GdnHCl. The fits shown in (b)–(d) were generated using the kinetic parameters tabulated in Table 1, which were obtained by carrying out global fitting of all kinetic traces at all GdnHCl concentrations at free chain concentrations of 10 μM.

Multiple folding pathways of monellin

The structure of dcMN is nearly identical with that of scMN, and the observation that their unfolding rates are similar[39] suggests that unfolding is dictated by the breaking of the same specific interactions in both variants. However, dcMN folds substantially slower than scMN.[7,42] The folding rate of a protein is correlated with its contact order,[72,73] because of the entropic costs of making long-range contacts between residues well separated along the sequence. It is possible that dcMN folds slower than scMN because the entropic costs of making long-range contacts would be considerably higher when the interacting residues are on different polypeptide chains. Although the two chains associate very rapidly during the folding reaction, the entropic cost of long-range interactions between residues on separate chains will still remain higher than if the interacting residues were on the same chain. It should also be noted that not only is the association of chains A and B dependent on protein concentration but that the subsequent observed folding/refolding rates of dcMN are also dependent on protein concentration because of kinetic coupling (see above). In contrast, the observed rate constants for the folding of scMN are independent of protein concentration. At a very high protein concentration, when the pre-equilibrium established between A, B, and C is such that virtually all chain molecules have formed the encounter complex C, the observed rate constants for the folding/refolding of dcMN are likely to exceed those of scMN.

Through the use of interrupted refolding experiments,[42] as well as of pulsed thiol labeling–mass spectrometry experiments,[7] it is known that scMN folds *via* multiple pathways. Although the fluorescence change accompanying the folding of scMN also occurs in three kinetic phases, all three folding chevron arms are linear. It should be noted that although an upward curvature in the folding chevron arm must mean that multiple folding pathways are operational, the converse need not be true. The upward curvature as a consequence of folding switching from one pathway to another will be seen only when the TSs on the pathways differ substantially in their compactness and structure. When the TSs have similar solvent-accessible surface areas, and hence similar interactions with denaturant, no curvature will be seen in the folding chevron arm.

Monellin belongs to the cystatin family of proteins, which all have the β-grasp fold.[74] The members of the cystatin family appear to fold by different folding mechanisms.[42,75,76] There are other examples of highly homologous proteins using different folding pathways.[77] Even circularly permuted variants of the same protein appear to fold by different pathways defined by TSs possessing different structures.[78,79] It is therefore not surprising that the folding mechanisms of scMN[42] and dcMN are different in their details, even though both variants utilize multiple pathways to fold. It has been suggested that folds displaying low TS structural conservation possess many possible folding pathways.[80] It remains to be seen if other members of the cystatin family also each fold *via* multiple pathways.

Materials and Methods

All buffers and reagents used in experiments were of ultrapure grade. GdnHCl was from USB Corporation. All the experiments were performed in 50 mM phosphate buffer containing 0.25 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM DTT at pH 7.

Purification of the dcMN

dcMN was purified as described previously.[46] Briefly, the two subunits of dcMN were expressed in *Escherichia coli* [BL21∗(DE3)] using a pET duet vector system. dcMN was purified using ion-exchange and gel-filtration chromatography. The purity of each protein was confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and mass spectrometry. The individual chains (A and B) of monellin were purified from dcMN by using reverse-phase chromatography. Separation was performed as described previously.[39]

Kinetic folding experiments

Experiments performed by mixing separated chains A and B are referred to as folding experiments. Folding studies were performed using an SFM-400 (Biologic) stopped-flow machine in the presence of different GdnHCl concentrations and at a fixed (10 μM) concentration of each chain. The dead time of mixing was 11.8 ms. Folding kinetic traces were monitored by measurement of the change in fluorescence at 340 ± 10 nm, using a band-pass filter (Asahi Spectra). The excitation wavelength was 280 nm, and the excitation slit width was 2 nm. Protein-concentration-dependent folding studies in the range of protein concentration from 2 to 30 μM were carried out by mixing equal concentrations of the separated chains to different final concentrations, in the absence of denaturant.

Kinetic refolding experiments

The protein was unfolded in unfolding buffer containing 2.5 M GdnHCl at pH 7 for at least 4 h. Refolding was initiated by dilution in native buffer to different final GdnHCl concentrations while maintaining the final protein concentration at 10 μM, using either SFM-400 stopped-flow machine (dead time of 11.8 ms) or manual mixing (dead time of 10 s).

Manual mixing experiments were for reactions with an observed rate constant less than 0.02 s^{-1}. Refolding was monitored as described above for the folding experiments. Protein-concentration-dependent refolding kinetic studies were performed by diluting unfolded protein to different

final protein concentrations and a fixed final GdnHCl concentration.

Double-jump experiment (interrupted refolding experiment)

Interrupted refolding experiments were carried out by manual mixing using Biologic MOS-450 optical system. dcMN (100 μM) was unfolded in 2.5 M GdnHCl and then refolded by diluting it to a final concentration of 10 μM in 0.25 M GdnHCl. The refolding reaction was interrupted at times ranging from 10 to 3000 s, by jumping the GdnHCl concentration to 5 M. The concentration of dcMN during subsequent unfolding was 2 μM. Resultant unfolding kinetic traces were acquired by monitoring the change in intrinsic tryptophan fluorescence at 340 nm, with the excitation wavelength set to 280 nm. The dead time of mixing was 10 s.

Data analysis

Kinetic folding and refolding traces

The kinetic data for folding were analyzed using a three-exponential equation, whereas those for refolding were analyzed using a two-exponential equation. Each kinetic trace at each GdnHCl concentration was individually fitted to obtain the observed rate constant (λ) and relative amplitude (α) of each of the observed kinetic phases of folding.

It should be noted that each λ and α value is determined by the forward and backward microscopic rate constants (k) of all the individual steps that define the folding mechanism. While analytical expressions can be obtained for the λ and α in terms of the microscopic rate constants, k, for simple two-state or three-state mechanisms,[81] such analytical expressions cannot be obtained for more complex mechanisms. Complex mechanisms such as those shown in Schemes 1–3 can be tested by computer simulations using numerical integration methods.

Double-jump experiment

The final unfolding kinetic traces in 5 M GdnHCl were analyzed using a single-exponential equation. The formation of the native state was fit to a single-exponential equation, while the formation and disappearance of the intermediate were fit to the equation:

$$Y_{(t)} = Y_0 + \left(\frac{\lambda_2}{\lambda_1 - \lambda_2}\right)\left(ae^{-(\lambda_1 t)} - be^{-(\lambda_2 t)}\right) \quad (1)$$

$Y_{(t)}$ is the observed signal at any time t; Y_0 is observed signal at $t = 0$; λ_1 and λ_2 are the observed rate constants for the formation and disappearance of the intermediate, respectively; and a and b represent amplitudes for each kinetic phase.

Chevron plot

Upward curvatures in the observed rate constants of refolding kinetics can be described empirically by the equation

$$\lambda_{obs} = \lambda_a e^{-\left(m_a^*[D]\right)} + \lambda_b e^{-\left(m_b^*[D]\right)} \quad (2)$$

λ_{obs} is the observed rate constant in water. The subscripts a and b represent two folding pathways, a and b, operating in parallel. On each pathway, λ is the rate constant for folding in the absence of denaturant, and m is the surface area change of the protein from U to the TS. [D] is the denaturant concentration.

Global fitting of kinetic data

Kinetic traces of folding and refolding at all GdnHCl concentrations were globally fit to Scheme 3 using a MATLAB program. Each rate constant in Scheme 3 was defined as: $\ln k_i = \ln k_i^0 + m_i$ [D], where m_i defines the exponential dependence of k_i on denaturant concentration. For fitting, the kinetic traces at different GdnHCl concentrations were simulated for Scheme 3, using the MATLAB function *ode23s*, using an initial set of values for the fitting parameters (k_i and m_i, see below). Simulations were carried out iteratively, and the fitting parameters were allowed to vary and optimized using the function *fminsearch* so as to achieve the lowest root-mean-square difference between all the experimental and all the simulated curves.

For global fitting, the bimolecular rate constant for the association of chains A and B was kept fixed to a value that ensured that chain association was fast compared to subsequent steps (see Discussion) at the concentrations of chains A and B used for the folding/refolding experiments. The three rate constants k_{-3a}, k_{-3b}, and k_{-4} were kept fixed at zero, because the folding reactions go to completion at the low denaturant concentrations used. It should be noted that in Scheme 3, only the association reaction of chains A and B to form complex C is dependent on free chain concentration, and the dependence is dictated by the value of the bimolecular rate constant k_1. All other reactions, including the dissociation of complex C to free chains A and B (defined by rate constant k_{-1}), are first-order reactions independent of protein concentration. For the different states defining Scheme 2, A, B, C, and N were assigned fluorescence values of 0, 0.43, 0.43, and 1, respectively, which are the experimental values after normalization of the native signal to 1. The fluorescence values for I_1 and I_2 were assumed to be independent of denaturant concentration and were allowed to be optimized along with the values of the rate constants during the global fitting process. Values of 0.7 and 0.9 were obtained for the fluorescence of I_1 and I_2, respectively.

Acknowledgements

We thank Shachi Gosavi and members of our laboratory for their comments on the manuscript. This work was funded by the Tata Institute of Fundamental Research and by the Department of Biotechnology, Government of India. J.B.U. is the recipient of a JC Bose National Fellowship from the Government of India.

References

1. Harrison, S. C. & Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl Acad. Sci. USA*, **82**, 4028–4030.
2. Ellison, P. A. & Cavagnero, S. (2006). Role of unfolded state heterogeneity and en-route ruggedness in protein folding kinetics. *Protein Sci.* **15**, 564–582.
3. Udgaonkar, J. B. (2008). Multiple routes and structural heterogeneity in protein folding. *Annu. Rev. Biophys.* **37**, 489–510.
4. Rami, B. R. & Udgaonkar, J. B. (2001). pH-jump-induced folding and unfolding studies of barstar: evidence for multiple folding and unfolding pathways. *Biochemistry*, **40**, 15267–15279.
5. Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **270**, 294–304.
6. Zaidi, F. N., Nath, U. & Udgaonkar, J. B. (1997). Multiple intermediates and transition states during protein unfolding. *Nat. Struct. Biol.* **4**, 1016–1024.
7. Jha, S. K., Dasgupta, A., Malhotra, P. & Udgaonkar, J. B. (2011). Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry. *Biochemistry*, **50**, 3062–3074.
8. Mallam, A. L. & Jackson, S. E. (2006). Probing nature's knots: the folding pathway of a knotted homodimeric protein. *J. Mol. Biol.* **359**, 1420–1436.
9. Mallam, A. L. & Jackson, S. E. (2007). A comparison of the folding of two knotted proteins: YbeA and YibK. *J. Mol. Biol.* **366**, 650–665.
10. Patra, A. K. & Udgaonkar, J. B. (2009). GroEL can unfold late intermediates populated on the folding pathways of monellin. *J. Mol. Biol.* **389**, 759–775.
11. Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. (2003). Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.* **10**, 658–662.
12. Sridevi, K. & Udgaonkar, J. B. (2002). Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry*, **41**, 1568–1578.
13. Jha, S. K. & Udgaonkar, J. B. (2010). Free energy barriers in protein folding and unfolding reactions. *Curr. Sci.* **99**, 457–475.
14. Sridevi, K., Juneja, J., Bhuyan, A. K., Krishnamoorthy, G. & Udgaonkar, J. B. (2000). The slow folding reaction of barstar: the core tryptophan region attains tight packing before substantial secondary and tertiary structure formation and final compaction of the polypeptide chain. *J. Mol. Biol.* **302**, 479–495.
15. Brooks, C. L., III, Gruebele, M., Onuchic, J. N. & Wolynes, P. G. (1998). Chemical physics of protein folding. *Proc. Natl Acad. Sci. USA*, **95**, 11037–11038.
16. Lakshmikanth, G. S., Sridevi, K., Krishnamoorthy, G. & Udgaonkar, J. B. (2001). Structure is lost incrementally during the unfolding of barstar. *Nat. Struct. Biol.* **8**, 799–804.
17. Chang, J. Y., Lu, B. Y. & Li, L. (2009). Fast and slow tracks in lysozyme folding elucidated by the technique of disulfide scrambling. *Protein J.* **28**, 300–304.
18. Nguyen, H., Jager, M., Moretto, A., Gruebele, M. & Kelly, J. W. (2003). Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl Acad. Sci. USA*, **100**, 3948–3953.
19. Sridevi, K., Lakshmikanth, G. S., Krishnamoorthy, G. & Udgaonkar, J. B. (2004). Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble. *J. Mol. Biol.* **337**, 699–711.
20. Yang, H. & Smith, D. L. (1997). Kinetics of cytochrome *c* folding examined by hydrogen exchange and mass spectrometry. *Biochemistry*, **36**, 14992–14999.
21. Dill, K. A. (1987). Protein surgery. *Protein Eng.* **1**, 369–371.
22. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19.
23. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science*, **267**, 1619–1620.
24. Bhuyan, A. K. & Udgaonkar, J. B. (1999). Observation of multistate kinetics during the slow folding and unfolding of barstar. *Biochemistry*, **38**, 9158–9168.
25. Goodsell, D. S. & Olson, A. J. (2000). Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153.
26. Rumfeldt, J. A., Galvagnion, C., Vassall, K. A. & Meiering, E. M. (2008). Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.* **98**, 61–84.
27. Sinha, K. K. & Udgaonkar, J. B. (2009). Early events in protein folding. *Curr. Sci.* **96**, 1053–1070.
28. Sosnick, T. R. & Barrick, D. (2011). The folding of single domain proteins—have we reached a consensus? *Curr. Opin. Struct. Biol.* **21**, 12–24.
29. Gloss, L. M. & Matthews, C. R. (1998). Mechanism of folding of the dimeric core domain of *Escherichia coli* trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway dimeric intermediate. *Biochemistry*, **37**, 15990–15999.
30. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L. & Matthews, C. R. (2000). Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.* **296**, 1105–1116.
31. Placek, B. J. & Gloss, L. M. (2005). Three-state kinetic folding mechanism of the H2A/H2B histone heterodimer: the N-terminal tails affect the transition state between a dimeric intermediate and the native dimer. *J. Mol. Biol.* **345**, 827–836.
32. Zhang, X., Kung, S. & Shan, S. O. (2008). Demonstration of a multistep mechanism for assembly of the SRP × SRP receptor complex: implications for the catalytic role of SRP RNA. *J. Mol. Biol.* **381**, 581–593.
33. Sugase, K., Dyson, H. J. & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, **447**, 1021–1025.
34. Mulder, A. M., Yoshioka, C., Beck, A. H., Bunner, A. E., Milligan, R. A., Potter, C. S. *et al.* (2010). Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science*, **330**, 673–677.
35. Ogata, C., Hatada, M., Tomlinson, G., Shin, W. C. & Kim, S. H. (1987). Crystal structure of the intensely sweet protein monellin. *Nature*, **328**, 739–742.
36. Somoza, J. R., Jiang, F., Tong, L., Kang, C. H., Cho, J. M. & Kim, S. H. (1993). Two crystal structures of a potently sweet protein. Natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution. *J. Mol. Biol.* **234**, 390–404.

37. Xue, W. F., Szczepankiewicz, O., Bauer, M. C., Thulin, E. & Linse, S. (2006). Intra- versus intermolecular interactions in monellin: contribution of surface charges to protein assembly. *J. Mol. Biol.* **358**, 1244–1255.
38. Tancredi, T., Iijima, H., Saviano, G., Amodeo, P. & Temussi, P. A. (1992). Structural determination of the active site of a sweet protein. A ^{1}H NMR investigation of pMNEI. *FEBS Lett.* **310**, 27–30.
39. Aghera, N., Earanna, N. & Udgaonkar, J. B. (2011). Equilibrium unfolding studies of monellin: the double-chain variant appears to be more stable than the single-chain variant. *Biochemistry*, **50**, 2434–2444.
40. Kimura, T., Maeda, A., Nishiguchi, S., Ishimori, K., Morishima, I., Konno, T. *et al.* (2008). Dehydration of main-chain amides in the final folding step of single-chain monellin revealed by time-resolved infrared spectroscopy. *Proc. Natl Acad. Sci. USA*, **105**, 13391–13396.
41. Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T. *et al.* (2005). Specific collapse followed by slow hydrogen- bond formation of beta-sheet in the folding of single-chain monellin. *Proc. Natl Acad. Sci. USA*, **102**, 2748–2753.
42. Patra, A. K. & Udgaonkar, J. B. (2007). Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways. *Biochemistry*, **46**, 11727–11743.
43. Jha, S. K. & Udgaonkar, J. B. (2009). Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc. Natl Acad. Sci. USA*, **106**, 12289–12294.
44. Jha, S. K., Dhar, D., Krishnamoorthy, G. & Udgaonkar, J. B. (2009). Continuous dissolution of structure during the unfolding of a small protein. *Proc. Natl Acad. Sci. USA*, **106**, 11113–11118.
45. Xue, W. F., Carey, J. & Linse, S. (2004). Multi-method global analysis of thermodynamics and kinetics in reconstitution of monellin. *Proteins*, **57**, 586–595.
46. Aghera, N. & Udgaonkar, J. B. (2011). Heterologous expression, purification and characterization of heterodimeric monellin. *Protein Expr. Purif.* **76**, 248–253.
47. Fan, P., Bracken, C. & Baum, J. (1993). Structural characterization of monellin in the alcohol-denatured state by NMR: evidence for beta-sheet to alpha-helix conversion. *Biochemistry*, **32**, 1573–1582.
48. Srivastava, A. K. & Sauer, R. T. (2000). Evidence for partial secondary structure formation in the transition state for arc repressor refolding and dimerization. *Biochemistry*, **39**, 8308–8314.
49. Gloss, L. M. & Matthews, C. R. (1998). The barriers in the bimolecular and unimolecular folding reactions of the dimeric core domain of *Escherichia coli* Trp repressor are dominated by enthalpic contributions. *Biochemistry*, **37**, 16000–16010.
50. Milla, M. E. & Sauer, R. T. (1994). P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry*, **33**, 1125–1133.
51. Mayr, L. M. & Schmid, F. X. (1993). Stabilization of a protein by guanidinium chloride. *Biochemistry*, **32**, 7994–7998.
52. Pradeep, L. & Udgaonkar, J. B. (2002). Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar. *J. Mol. Biol.* **324**, 331–347.
53. Pradeep, L. & Udgaonkar, J. B. (2004). Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Biol. Chem.* **279**, 40303–40313.
54. Schreiber, G. & Fersht, A. R. (1993). The refolding of *cis*- and *trans*-peptidylprolyl isomers of barstar. *Biochemistry*, **32**, 11195–11203.
55. Shastry, M. C., Agashe, V. R. & Udgaonkar, J. B. (1994). Quantitative analysis of the kinetics of denaturation and renaturation of barstar in the folding transition zone. *Protein Sci.* **3**, 1409–1417.
56. Agashe, V. R. & Udgaonkar, J. B. (1995). Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry*, **34**, 3286–3299.
57. Greene, R. F., Jr & Pace, C. N. (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393.
58. Tanford, C. (1968). Protein denaturation. *Adv. Protein Chem.* **23**, 121–282.
59. Shastry, M. C. & Udgaonkar, J. B. (1995). The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* **247**, 1013–1027.
60. Park, S. H., Shastry, M. C. & Roder, H. (1999). Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat. Struct. Biol.* **6**, 943–947.
61. Juneja, J. & Udgaonkar, J. B. (2002). Characterization of the unfolding of ribonuclease A by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding. *Biochemistry*, **41**, 2641–2654.
62. Sanchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.
63. Oliveberg, M., Tan, Y. J., Silow, M. & Fersht, A. R. (1998). The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J. Mol. Biol.* **277**, 933–943.
64. Silow, M. & Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl Acad. Sci. USA*, **94**, 6084–6086.
65. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.* **324**, 359–371.
66. Bhuyan, A. K. & Udgaonkar, J. B. (2001). Folding of horse cytochrome c in the reduced state. *J. Mol. Biol.* **312**, 1135–1160.
67. Zhang, Z. & Chan, H. S. (2010). Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proc. Natl Acad. Sci. USA*, **107**, 2920–2925.
68. Chan, H. S., Zhang, Z., Wallin, S. & Liu, Z. (2011). Cooperativity, local–nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.* **62**, 301–326.
69. Fersht, A. R., Itzhaki, L. S., elMasry, N. F., Matthews, J. M. & Otzen, D. E. (1994). Single versus parallel

pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl Acad. Sci. USA*, **91**, 10426–10429.
70. Fowler, S. B. & Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure*, **9**, 355–366.
71. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* **337**, 463–475.
72. Jelinska, C., Davis, P. J., Kenig, M., Zerovnik, E., Kokalj, S. J., Guncar, G. *et al.* (2011). Modulation of contact order effects in the two-state folding of stefins A and B. *Biophys. J.* **100**, 2268–2274.
73. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
74. Murzin, A. G. (1993). Sweet-tasting protein monellin is related to the cystatin family of thiol proteinase inhibitors. *J. Mol. Biol.* **230**, 689–694.
75. Krantz, B. A. & Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry*, **39**, 11696–11701.
76. Zerovnik, E., Jerala, R., Virden, R., Kroon Zitko, L., Turk, V. & Waltho, J. P. (1998). On the mechanism of human stefin B folding: II. Folding from GuHCl unfolded, TFE denatured, acid denatured, and acid intermediate states. *Proteins*, **32**, 304–313.
77. Nakamura, T., Makabe, K., Tomoyori, K., Maki, K., Mukaiyama, A. & Kuwajima, K. (2010). Different folding pathways taken by highly homologous proteins, goat alpha-lactalbumin and canine milk lysozyme. *J. Mol. Biol.* **396**, 1361–1378.
78. Lindberg, M., Tangrot, J. & Oliveberg, M. (2002). Complete change of the protein folding transition state upon circular permutation. *Nat. Struct. Biol.* **9**, 818–822.
79. Viguera, A. R., Serrano, L. & Wilmanns, M. (1996). Different folding transition states may result in the same native structure. *Nat. Struct. Biol.* **3**, 874–880.
80. Zarrine-Afsar, A., Larson, S. M. & Davidson, A. R. (2005). The family feud: do proteins with similar structures fold via the same pathway? *Curr. Opin. Struct. Biol.* **15**, 42–49.
81. Szabo, Z. G. (1969). Kinetic characterization of complex reaction systems. In *Comprehensive Chemical Kinetics* (Bamford, C. H. T. & Tipper, C. F. H., eds), 2, pp. 1–80. Elsevier Publishing Company, Amsterdam, The Netherlands.