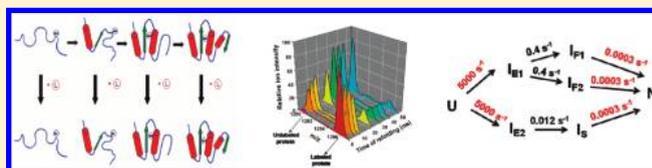


# Identification of Multiple Folding Pathways of Monellin Using Pulsed Thiol Labeling and Mass Spectrometry

Santosh Kumar Jha, Amrita Dasgupta, Pooja Malhotra, and Jayant B. Udgaonkar\*

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

**ABSTRACT:** Protein folding reactions often display multi-exponential kinetics of changes in intrinsic optical signals, as a manifestation of heterogeneity, either on one folding pathway or on multiple folding pathways. Delineating the origin of this heterogeneity is difficult because different coexisting structural forms of a protein cannot be easily distinguished by optical probes. In this study, the complex folding reaction of single-chain monellin has been investigated using a pulsed thiol labeling (SX) methodology in conjunction with mass spectrometry, which measures the kinetics of burial of a cysteine side chain thiol during folding. Because it can directly distinguish between unfolded and folded molecules and can measure the disappearance of the former during folding, the pulsed SX methodology is an ideal method for investigating whether multiple pathways are operative during folding. The kinetics of burial of the C42 thiol of monellin was observed to follow biexponential kinetics. To determine whether this was because the fast phase leads to the partial protection of the thiol group in all the molecules or to complete protection in only a fraction of the molecules, the duration and intensity of the labeling pulse were varied. The observation that the extent of labeling did not vary with the duration of the pulse cannot be explained by a simple sequential folding mechanism. Two parallel folding pathways are shown to be operative, with one leading to the formation of thiol-protective structure more rapidly than the other.



The classical view of protein folding is that all unfolded protein molecules fold via essentially the same sequence of structural events to reach the native state.<sup>1–5</sup> A different view, based upon statistical mechanical models, is that folding occurs via a multitude of routes traversing a large energy landscape.<sup>6–14</sup> In this view, intermediate states en route to folding are also visualized as ensembles of different structural forms whose members can follow different folding trajectories on parallel routes. Although it is relatively easy to show the presence of parallel and multiple folding pathways in computer simulations, it has remained challenging to do so in experimental studies.<sup>15–32</sup> This has been partly because most of the techniques that are commonly used to probe folding reactions measure ensemble-averaged changes in signals and do not directly measure the changes in the fractional populations of unfolded (U), intermediate (I), and native (N) conformations during folding.

The pulsed thiol labeling (SX) methodology provides direct structural information about the fates of individual residue side chains during the folding process and has been shown to be an excellent probe for studying structure formation and/or dissolution during the folding–unfolding reactions of several proteins.<sup>33–37</sup> A library of single cysteine-containing mutant proteins is first created, in each of which the cysteine thiol is located in a different part of the protein structure. At different times of folding, the extent to which a particular cysteine thiol group is labeled by a short pulse of thiol labeling reagent is measured. The extent of labeling reflects the fraction of molecules in which the cysteine thiol is not protected by structure that has already formed at the time of application of the labeling pulse. The pulsed SX methodology is an ideal method for investigating

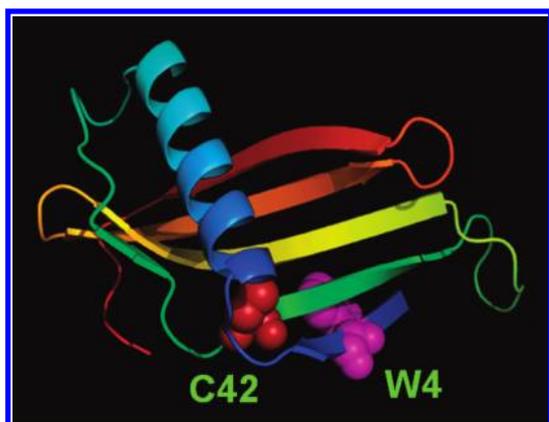
whether multiple pathways are operative during the folding of a protein, because it can measure directly the disappearance of unfolded molecules during folding. In this study, the pulsed SX methodology has been coupled with mass spectrometry and has been used to study the folding of single-chain monellin (MNEI), an intensely sweet, small plant protein (Figure 1).

The folding and unfolding of MNEI has been characterized in detail in earlier studies, using optical probes,<sup>30,38–40</sup> and the mechanism of folding was found to be complex. Folding was shown to occur in five well-separated kinetic phases: ultrafast ( $\sim 5000 \text{ s}^{-1}$ ), very fast ( $\sim 10 \text{ s}^{-1}$ ), fast ( $\sim 1 \text{ s}^{-1}$ ), slow ( $\sim 0.1 \text{ s}^{-1}$ ), and very slow ( $\sim 0.002 \text{ s}^{-1}$ ). Intrinsic tryptophan fluorescence was found to change only during the very fast, fast, and slow kinetic phases of folding. Nevertheless, the ultrafast and very slow kinetic phases could be monitored by measurement of the binding of the extrinsic fluorophore ANS, whose fluorescence increases during the ultrafast phase<sup>38</sup> and decreases during the very fast, fast, and very slow phases.<sup>30</sup> The ultrafast phase was found to lead to the formation of intermediate ensemble  $I_E$ , and it was proposed that  $I_E$  consists of at least two subpopulations,  $I_{E1}$  and  $I_{E2}$ . A double-jump, interrupted folding assay could monitor the formation of the relative amounts of differently folded forms on the basis of how fast they unfold. It showed that only the very slow kinetic phase leads to the formation of N, while the other kinetic phases lead to the formation of partially folded intermediates,<sup>30</sup>  $I_{VF}$  from the very fast phase,  $I_F$  from the fast

Received: April 24, 2010

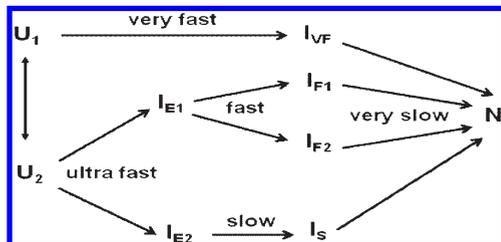
Revised: January 7, 2011

Published: March 01, 2011



**Figure 1.** Structure of single-chain monellin (MNEI). The locations of C42 in the third  $\beta$ -strand in the core and W4 in the first  $\beta$ -strand are shown. The structure was drawn from Protein Data Bank entry 1IV7 using PyMOL.<sup>62</sup>

**Scheme 1**



phase, and  $I_S$  from the slow phase. The double-jump assay showed that  $I_F$  comprises two subpopulations,  $I_{F1}$  and  $I_{F2}$ . A double-jump, interrupted unfolding assay could monitor the formation of the relative amounts of two unfolded forms,  $U_1$  and  $U_2$ , on the basis of how they refold. It showed that  $U_1$  gives rise to the very fast phase of folding while  $U_2$  gives rise to both the fast and slow phases of folding. Scheme 1 was found to adequately and minimally describe all the data for folding at pH 7 and 25 °C.<sup>30</sup>

Hence, the optical and double-jump experiments have shown that MNEI folds via multiple pathways, which are delineated by intermediates that form on very different time scales. It is therefore an ideal model protein for demonstrating the efficacy of the pulsed SX methodology in delineating, directly and quantitatively, the existence of multiple pathways, in showing directly the existence of very early intermediates, and in determining the relative stabilities of the different intermediates. Moreover, MNEI contains only a single cysteine residue (C42) in its sequence (Figure 1) and, hence, offers itself as an amenable model system for pulsed SX studies.

In this study, the kinetics of refolding of MNEI was probed by measuring the change in intrinsic tryptophan fluorescence as well as the change in the solvent accessibility of the sole cysteine thiol. The solvent accessibility of the single cysteine thiol has been measured at different times during the folding process, by determining the extent to which the thiol is labeled by a 4 ms pulse of methyl methanethiosulfonate (MMTS), which is known to label solvent-exposed thiols very rapidly, with a bimolecular rate constant of  $\sim 5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>34,37,41</sup> It is directly shown

that unfolded molecules disappear on parallel routes during the refolding of this protein.

## MATERIALS AND METHODS

**Protein Expression and Purification.** MNEI was expressed and purified as reported previously.<sup>30</sup> The purity of the protein was checked by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and was found to be >95%. The mass of the protein was confirmed to be 11403 Da by mass spectrometry using a Micromass Q-TOF Ultima mass spectrometer coupled with an ESI source. Protein concentrations were determined by measurement of the absorbance at 280 nm, using an extinction coefficient of  $14600 \text{ M}^{-1} \text{ cm}^{-1}$ .<sup>30</sup>

**Reagents and Chemicals.** Glycine, EDTA disodium salt, MMTS, 5,5'-dithiobis(2-nitrobenzoic acid), and cysteine hydrochloride were of ultrapure grade from Sigma. GdnHCl (ultrapure grade) was from GE. Dithiothreitol (ultrapure grade) was obtained from Invitrogen; formic acid (GPR grade) was from BDH or Sigma (MS grade), and acetonitrile (HPLC grade) was from Qualigens or Sigma (MS grade).

**Buffers and Solutions.** The native buffer used in all the experiments consisted of 50 mM glycine and 1 mM EDTA (pH  $9.4 \pm 0.1$ ). The unfolding buffer was native buffer containing 6 M GdnHCl. The concentration of a GdnHCl solution was determined by measurement of the refractive index on an Abbe 3L refractometer from Milton Roy. All buffers and solutions were filtered through  $0.22 \mu\text{m}$  filters and degassed before being used. All of the experiments were conducted at 25 °C.

**Fluorescence Spectra.** The fluorescence spectra were recorded on a Spex Fluoromax 3 spectrofluorimeter as described previously,<sup>30</sup> with the excitation wavelength set to 280 nm. The protein concentration was  $\sim 5 \mu\text{M}$ .

**Fluorescence-Monitored Equilibrium and Kinetic Folding Experiments.** For equilibrium unfolding experiments, the native protein was incubated in different concentrations of GdnHCl ranging from 0 to 6 M for 6 h, and the fluorescence signals were measured on the stopped-flow module (SFM-4) from Biologic. Sample excitation was conducted at 280 nm, and emission was monitored at 340 nm by using a band-pass filter.

The kinetics of folding of MNEI was measured on the stopped-flow mixing module (SFM-4) from Biologic. The protein was unfolded in 3 M GdnHCl (unfolding buffer) for at least 6 h prior to the refolding experiments. Refolding was initiated by appropriate dilution of refolding buffer, unfolding buffer, and unfolded protein inside the stopped-flow mixing module, to give the final desired GdnHCl concentration at the time of refolding. Sample excitation was conducted at 280 nm, and emission was monitored at 340 nm by using a band-pass filter. In all the experiments, a mixing dead time of 1.8 ms was achieved by using an FC-08 cuvette with a path length of 0.8 mm and a total flow rate of 5 mL/s. The final protein concentration used in the fluorescence-monitored kinetic refolding experiments was 20–25  $\mu\text{M}$ .

The folding kinetics of MNEI was also measured following a pH jump from 12 to 9.4. The protein was unfolded at pH 12 in water (pH adjusted to 12 with NaOH). The refolding reaction was initiated by mixing the unfolded protein with refolding buffer [60 mM glycine and 1 mM EDTA (pH 9.4)] containing the desired amount of GdnHCl, in a 1:5 ratio, and the change in fluorescence was monitored as described above.

**Preparation of MMTS-Labeled Protein.** MMTS-labeled protein was prepared as described previously.<sup>37</sup> In brief, MNEI was unfolded in 6 M GdnHCl and the unfolded protein was incubated with a 100-fold molar excess of MMTS for ~5 min at pH 9.4. The labeling reaction was quenched by the addition of a 1000-fold molar excess (compared to the protein concentration) of cysteine hydrochloride (in 1% formic acid) to the reaction mixture. The final pH after quenching was ~2, which ensured that no label came out of the labeled protein. The MMTS-labeled protein was separated from cysteine, GdnHCl, and other small molecules present in the reaction mixture by being passed through a Hi-Trap Sephadex G-25 desalting column using an Akta chromatography system. The extent of labeling was checked by mass spectrometry, and the protein was found to be fully labeled as judged by the observed 46 Da increase in its mass upon labeling, and the absence of any peak in the mass spectrum corresponding to the mass of the unlabeled protein.

**Kinetics of the Change in Cysteine Accessibility during Refolding.** All pulsed SX experiments were conducted using a Biologic SFM-400 Q/S unit operating in the quenched-flow mode. Both refolding and labeling reactions were performed at pH 9.4 and 25 °C. The protein was unfolded in 3 M GdnHCl (unfolding buffer) for at least 3 h prior to the refolding experiments. At different times of refolding, a 4 ms pulse of MMTS label was applied. The labeling reaction was quenched by the addition of excess cysteine in 1% formic acid.

The kinetics of the change in the protection of the cysteine thiol was measured for folding in 0.3 and 0.7 M GdnHCl. Two different quenched-flow programs were used. For labeling at the 0 ms refolding time point during folding in 0.3 M GdnHCl, 30  $\mu$ L of 26.7 mM MMTS (in water) was mixed with 330  $\mu$ L of refolding buffer inside the quenched-flow machine for 5 ms, and the resulting solution was mixed with 40  $\mu$ L of unfolded protein solution (150  $\mu$ M stock unfolded in 3 M GdnHCl) for 4 ms. The labeling reaction was quenched via the addition of 103  $\mu$ L of 77.7 mM cysteine hydrochloride (in 1% formic acid). For labeling at refolding times greater than 0 ms during folding in 0.3 M GdnHCl, refolding was initiated by mixing 21  $\mu$ L of the unfolded protein solution (225  $\mu$ M stock unfolded in 3 M GdnHCl) with 195  $\mu$ L of refolding buffer, at pH 9.4 and 25 °C in a 190  $\mu$ L delay loop (total intermixer volume of 216  $\mu$ L). After a variable delay time, a pulse consisting of 100  $\mu$ L of 6.32 mM MMTS (in water) was applied for 4 ms. The labeling reaction was quenched via the addition of 100  $\mu$ L of 63.2 mM cysteine hydrochloride (in 1% formic acid).

For labeling at the 0 ms refolding time point during folding in 0.7 M GdnHCl, 30  $\mu$ L of 26.7 mM MMTS (in water) was mixed with 330  $\mu$ L of refolding buffer containing 0.49 M GdnHCl, inside the quenched-flow machine for 5 ms, and the resulting solution was mixed with 40  $\mu$ L of unfolded protein solution (150  $\mu$ M stock unfolded in 3 M GdnHCl) for 4 ms. The labeling reaction was quenched via the addition of 103  $\mu$ L of 77.7 mM cysteine hydrochloride (in 1% formic acid). For labeling at refolding times greater than 0 ms during folding in 0.7 M GdnHCl, refolding was initiated by mixing 21  $\mu$ L of the unfolded protein solution (225  $\mu$ M stock unfolded in 3 M GdnHCl) with 195  $\mu$ L of refolding buffer containing 0.45 M GdnHCl, at pH 9.4 and 25 °C in a 190  $\mu$ L delay loop (total intermixer volume of 216  $\mu$ L). After a variable delay time, a pulse consisting of 100  $\mu$ L of 6.32 mM MMTS (in water) was applied for 4 ms. The labeling reaction was quenched via the addition of 100  $\mu$ L of 63.2 mM cysteine hydrochloride (in 1% formic acid).

The concentration of the protein at the time of labeling was 15  $\mu$ M in all the pulsed SX experiments. The concentration of MMTS at the time of labeling was 2 mM, except in the experiments where the dependence on MMTS concentration of the extent of labeling of the thiol, in the fully unfolded protein as well as in the native protein, was measured. In those experiments, unfolded protein was mixed with unfolding buffer (and the native protein was mixed with refolding buffer) and pulsed with the calculated amount of the MMTS solution (in water), to give the desired MMTS concentration at the time of labeling. The concentration of cysteine hydrochloride in all of the experiments described above was 10-fold higher than that of the MMTS at the time of quenching.

Control experiments were conducted to ensure that the concentrations of MMTS and cysteine used in the experiments described above were sufficient to fully label the unfolded protein and quench the labeling reaction, respectively. All of the pulsed SX experiments were completed within 2 h of the preparation of the MMTS solution.<sup>37</sup>

**Processing of Samples for Mass Spectrometry.** All samples from the pulsed SX experiments were processed in an identical manner. Each sample was desalted on an Akta chromatography system, using a Hi-Trap Sephadex G-25 desalting column. Milli-Q water at pH 3 (pH adjusted with formic acid) was used for elution. Control experiments were performed, as discussed previously,<sup>37</sup> to ensure that there was no cross-contamination between two samples during desalting, and that the labeled protein was not reduced due to the presence of a high concentration of free cysteine in the samples after the pulsed SX experiments. All the samples were desalted within 2 h of completion of the pulsed SX experiments.

**Determination of the Extent of Labeling by ESI Mass Spectrometry.** The extents of labeling in the samples from the pulsed SX experiments were determined using ESI mass spectrometry as described previously.<sup>37</sup> Either a Q-TOF Ultima or a SYNAPT G2 mass spectrometer from Waters Corp., coupled with an ESI source and operated under Mass Lynx software control, was used. For acquisition of the mass spectra on the Q-TOF Ultima mass spectrometer, the capillary and cone voltages were maintained at 3 kV and 80 V, respectively, the desolvation temperature was set to 150 °C, and the source temperature was set to 80 °C. For acquisition of mass spectra on the Synapt G2 mass spectrometer, the capillary and cone voltages were maintained at 2.76 kV and 80 V, respectively, the desolvation temperature was set to 200 °C, and the source temperature was set to 80 °C. Samples collected after desalting were mixed with acetonitrile (containing 0.2% formic acid) in a 1:1 ratio and were infused into the mass spectrometer using a syringe pump (Harvard Apparatus, Holliston, MA) at a flow rate of 10  $\mu$ L/min (Q-TOF Ultima) or 20  $\mu$ L/min (Synapt G2). All of the spectra were recorded in positive ion mode. The concentration of the protein in each sample was typically 2–3  $\mu$ M, and typically, an ion count of at least ~125 was obtained in a 1 s data acquisition window. Instrument calibration was achieved with a separate injection of horse heart myoglobin.

Typically, a mass spectrum consisting of a series of multiply charged peaks corresponding to the masses of the unlabeled and MMTS-labeled proteins was observed in each 1 s scan. For each sample, the data acquired over 100 s were averaged. All of the resultant *m/z* spectra were processed in the following way using Mass Lynx version 4.0 or 4.1. Background noise was subtracted using a second-order polynomial below 30% of the curve with a

tolerance value of 0.01, followed by a two-point smoothing with a Savitzky–Golay algorithm (supplied with Mass Lynx) using a smoothing window (in channels) of  $\pm 23$ . The extent of labeling was determined from these smoothed  $m/z$  spectra via calculation of the average relative ion intensity of the labeled protein from the 10th, 11th, 12th, and 13th charged state peaks (these were the four most intense peaks in the mass spectra).

**Data Analysis.** *Fluorescence-Monitored Kinetic Data.* In each case, typically six or seven kinetic traces were averaged, and the resultant traces were fitted to the sum of two exponentials using the equation

$$A(t) = A_0 + A_1(1 - e^{-\lambda_1 t}) + A_2(1 - e^{-\lambda_2 t}) \quad (1)$$

where  $A(t)$  and  $A_0$  are the observed amplitudes at time  $t$  and time zero, respectively,  $\lambda_1$  and  $\lambda_2$  are the apparent rate constants of the fast and slow phases, respectively, and  $A_1$  and  $A_2$  are the respective amplitudes. The relative amplitudes of the fast and slow phases were determined using eqs 2 and 3, respectively:

$$\alpha_1 = \frac{A_1}{S_U - S_N} \times 100 \quad (2)$$

$$\alpha_2 = \frac{A_2}{S_U - S_N} \times 100 \quad (3)$$

where  $\alpha_1$  and  $\alpha_2$  are the relative amplitudes of the fast and slow phases, respectively, and  $S_U$  and  $S_N$  are the fluorescence signals of the unfolded protein and native protein, respectively.

*Cysteine Burial-Monitored Kinetic Data.* The cysteine accessibility-monitored refolding kinetics was fitted to the double-exponential decay equation

$$A'(t) = A'_0 + A'_1 e^{-\lambda_3 t} + A'_2 e^{-\lambda_4 t} \quad (4)$$

where  $A'(t)$  and  $A'_0$  are the observed amplitudes at time  $t$  and time zero, respectively,  $\lambda_3$  and  $\lambda_4$  are the apparent rate constants of the fast and slow phases, respectively, and  $A'_1$  and  $A'_2$  are the respective amplitudes. The relative amplitudes of the fast and slow phases were determined using eqs 5 and 6, respectively:

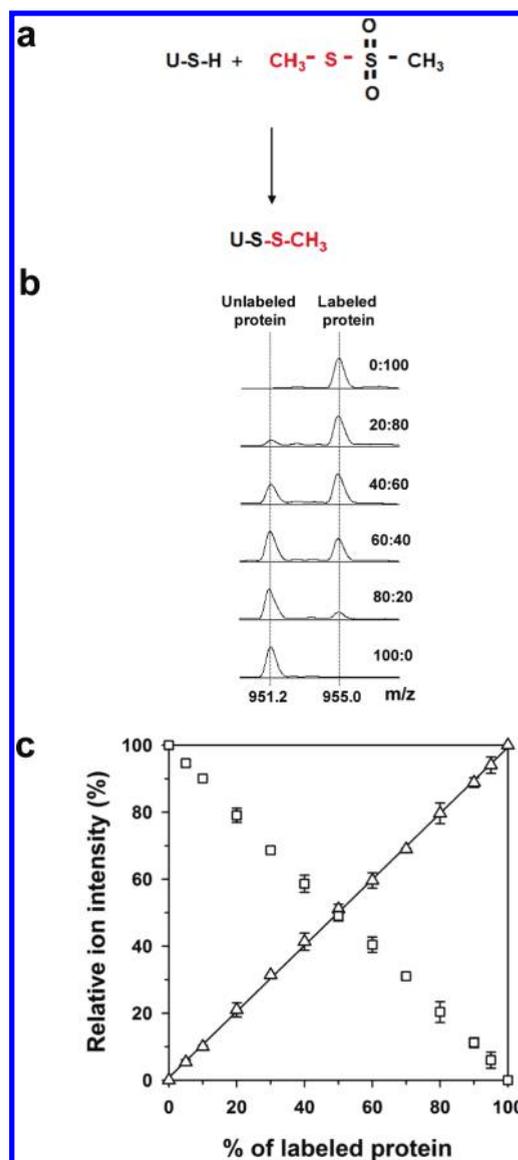
$$\alpha'_1 = \frac{A'_1}{S'_U - S'_N} \times 100 \quad (5)$$

$$\alpha'_2 = \frac{A'_2}{S'_U - S'_N} \times 100 \quad (6)$$

where  $\alpha'_1$  and  $\alpha'_2$  are the relative amplitudes of the fast and slow phases, respectively, and  $S'_U$  and  $S'_N$  are the extents of labeling of the unfolded protein and native protein, respectively.

## RESULTS

In this study, measurements of the change in the intrinsic fluorescence of W4, as well as of the change in the solvent accessibility of the thiol side chain of C42, have been used to monitor the folding of MNEI. The fluorescence measurements report on changes in the environment of W4. The change in the accessibility of the thiol side chain of C42 during folding has been monitored by determining the extent to which the thiol is labeled by the thiol labeling reagent MMTS at different times of folding. The extent of labeling at any time of folding has been determined by ESI mass spectrometry. Before the application of the pulsed SX methodology in studying the folding of MNEI, it was first necessary to standardize folding and labeling conditions under

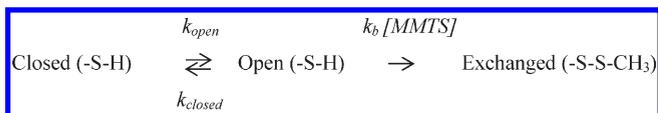


**Figure 2.** Quantitative estimation of the extent of the labeling of the cysteine thiol in MNEI. (a) MMTS reacts with a solvent-exposed thiol group of a protein, increasing the mass of the protein by 46 Da. (b) Mass spectra (12th charged state) of mixtures of unlabeled and MMTS-labeled MNEI, which had been mixed at the indicated molar ratios. (c) Relative populations of unlabeled ( $\square$ ) and fully labeled ( $\triangle$ ) MNEI in a mixture of the two, determined from their relative ion intensities (calculated using the average of the 10th, 11th, 12th, and 13th charged states) in the mass spectra, scale linearly with the molar ratio at which the labeled and unlabeled molecules were mixed. In panel c, the error bars represent the standard deviations from three separate experiments.

which all unfolded protein molecules would be fully labeled but fully folded protein molecules would not be labeled at all by the labeling pulse of MMTS. It was also necessary to standardize data acquisition parameters on the mass spectrometer, so that the signal intensity was proportional to protein concentration over a defined range of protein concentrations.

**ESI-MS Can Determine Quantitatively the Relative Amounts of Labeled and Unlabeled Proteins in a Mixture of the Two.** Figure 2a shows the chemical reaction between a solvent-exposed cysteine thiol group of a protein and MMTS.

## Scheme 2



MMTS transfers one -S-CH<sub>3</sub> moiety to the thiol group and increases the mass of the protein by 46 Da. An ESI mass spectrum of a sample consisting of a mixture of labeled and unlabeled MNEI shows two peaks, 46 Da apart in absolute mass, corresponding to the masses of the two proteins. To establish that it is possible to determine the relative amount of labeled and unlabeled proteins present in a mixture of the two, labeled and unlabeled proteins were mixed in different known ratios and fed into the mass spectrometer. Figure 2b shows the mass spectra (12th charged state) of six samples in which unlabeled and MMTS-labeled MNEI were mixed at the indicated molar ratio. As one can see from these mass spectra, the relative peak intensities of the unlabeled and labeled proteins are found to be directly proportional to their relative amounts in the mixture. In Figure 2c, it is shown that the relative ion intensities of labeled and unlabeled proteins, as determined from the average of the 10th, 11th, 12th, and 13th charged states of the mass spectrum, predict the correct molar ratio in which the two proteins were mixed. Hence, ESI-MS can be used to determine quantitatively the relative amounts of labeled and unlabeled proteins in a mixture of the two.

#### Establishing the Conditions for Pulsed Cysteine Labeling.

The exchange reaction between a thiol labeling reagent and a protected thiol group of a protein is similar to amide–hydrogen exchange reactions in proteins<sup>42</sup> and can be modeled by a Linderstrom–Lang type of mechanism<sup>34,36,37</sup> as follows:

A cysteine thiol protected in the protein structure can be labeled with MMTS only when a structural opening reaction, i.e., local or global unfolding, exposes that thiol transiently to the solvent. In Scheme 22,  $k_{\text{open}}$  and  $k_{\text{closed}}$  are the rate constants for opening and closing, respectively, of a cysteine thiol residue and  $k_b$  is the second-order rate constant of the reaction of that thiol group with MMTS in the unfolded protein. The observed rate constant of exchange of the thiol in the closed state is given by

$$k_{\text{ex}} = \frac{k_{\text{open}}k_b[\text{MMTS}]}{k_{\text{closed}} + k_b[\text{MMTS}]} \quad (7)$$

Depending upon the relative rates of the closing reaction and of chemical exchange from the open unfolded state ( $k_b[\text{MMTS}]$ ), two limiting cases for eq 7 exist.

If  $k_{\text{closed}} \ll k_b[\text{MMTS}]$ , then

$$k_{\text{ex}} = k_{\text{open}} \quad (8)$$

Under this condition,  $k_{\text{ex}}$  measures the rate of structural opening in the closed-to-open reaction. This is known as the SX1 limit.

If  $k_{\text{closed}} \gg k_b[\text{MMTS}]$ , then

$$k_{\text{ex}} = \frac{k_{\text{open}}k_b[\text{MMTS}]}{k_{\text{closed}}} = K_{\text{open}}k_b[\text{MMTS}] \quad (9)$$

Under this condition,  $k_{\text{ex}}$  measures the equilibrium constant ( $K_{\text{open}}$ ) between closed and open states, and this is known as the SX2 limit. In this limit, the rate of labeling of the thiol depends

upon the concentration of MMTS (eq 9), and hence, it can be readily distinguished from the SX1 limit (see eq 8).

It should be noted that the pulsed SX labeling experiments are kinetic experiments. If there are unstructured protein molecules present with their thiol group unprotected at the time of application of the labeling pulse, the thiol group in these molecules will be labeled at the rate  $k_b[\text{MMTS}]$ . Because the experiment is designed in such a manner that the duration of the pulse is short compared to the folding time, labeling of a protected thiol group in structured protein molecules will occur under essentially pseudoequilibrium conditions. Scheme 22 will then be applicable for all protein molecules that have become partially or fully structured at the time of application of the labeling pulse, because for these molecules, a structural opening reaction that exposes the thiol group to solvent must occur before labeling can occur.

If a single population of molecules lacking any protection against exchange is present at the time ( $t$ ) of application of a labeling pulse of MMTS of duration  $t_p$  (short in comparison to the folding time  $t$ ), then using eq 9, the fraction ( $f$ ) of molecules that will be labeled is given by

$$f = 1 - e^{-k_b[\text{MMTS}]t_p} \quad (10)$$

On the other hand, if a single population of protected molecules is present, then

$$f = 1 - e^{-k_b[\text{MMTS}]K_{\text{open}}t_p} \quad (11)$$

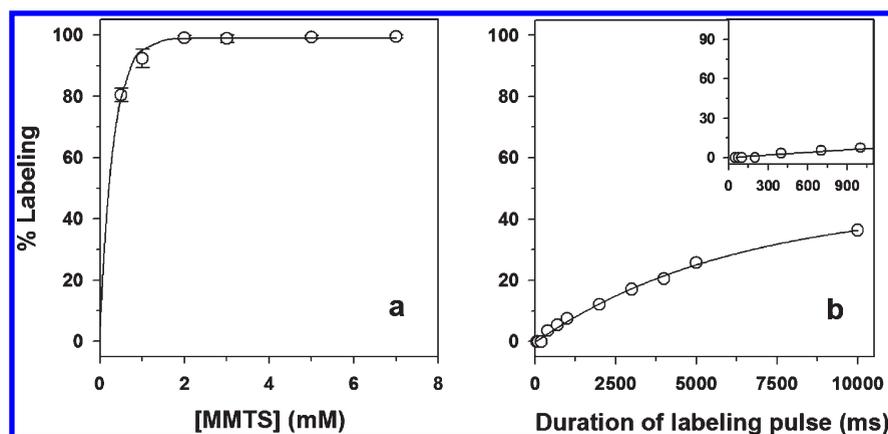
If two populations of molecules, one unprotected and the other protected, are present at the time of application of the labeling pulse of MMTS, then

$$f = f_U(1 - e^{-k_b[\text{MMTS}]t_p}) + (1 - f_U)(1 - e^{-k_b[\text{MMTS}]K_{\text{open}}t_p}) \quad (12)$$

where  $f_U$  is the fraction of molecules that are unprotected at folding time  $t$ .

In the pulsed SX experiments, it was necessary to ensure that the labeling pulse fully labels a thiol exposed fully to the solvent, i.e., in the unfolded state. To determine the strength of the labeling pulse to be used for pulsed SX experiments with MNEI, unfolded MNEI was incubated with different concentrations of MMTS (ranging from 0.5 to 7 mM) for 4 ms, inside the quenched-flow module, and the extent of labeling was determined by mass spectrometry (Figure 3a). When the data in Figure 3a were fit to eq 10, a value for  $k_b$  of  $8 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  was obtained. This value is similar to the value of  $4.8 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  obtained for  $k_b$  for the reaction between MMTS and a thiol fully exposed to solvent in urea-unfolded barstar.<sup>37</sup> The data show that a 4 ms labeling pulse of  $\geq 2$  mM MMTS will fully label the thiol in GdnHCl-unfolded MNEI (Figure 3a).

It was also important to ensure that the labeling pulse did not label the thiol in the native protein. To determine the extent to which the cysteine thiol is protected in the native protein, we incubated native MNEI with 2 mM MMTS (the lowest concentration that is sufficient to label the thiol in unfolded protein completely) for different durations of time. Figure 3b shows that native MNEI is not labeled at all even when a 2 mM pulse of MMTS is applied for  $\sim 200$  ms and that it is labeled in only  $\sim 40\%$  of the molecules when the duration of the pulse is increased to 10 s. This result suggested that the cysteine thiol is at least 10000



**Figure 3.** Determination of the accessibility of C42 to MMTS labeling in (a) unfolded protein and (b) native protein. In panel a, the extent of labeling in unfolded MNEI (in 3 M GdnHCl) is plotted vs the concentration of MMTS present during the 4 ms labeling pulse. The solid line through the data is a nonlinear, least-squares fit of the data to eq 10 and yields a value for  $k_b$  of  $8 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ . In panel b, the extent of labeling in native MNEI (in 0 M GdnHCl) is plotted vs the duration of the labeling pulse containing 2 mM MMTS. The inset shows the first 1100 ms of the data. The solid lines through the data are drawn only to help visual inspection.

times more protected in native MNEI than in the unfolded protein [ $P_{\text{SX}} = 1/K_{\text{open}}$  (see eq 9)].

**Characterization of the Refolding Reaction of MNEI by Measurement of the Change in Fluorescence.** Before using the change in solvent accessibility of the thiol as a probe to study refolding, it was important to use the more traditional fluorescence measurements to obtain a basic description of the folding of MNEI at pH 9.4 and 25 °C. Figure 4a shows that the wavelength of maximal fluorescence emission of native MNEI is 345 nm, whereas that of the unfolded protein is 356 nm. MNEI contains seven tyrosine residues and one tryptophan residue in its sequence. Upon excitation at 280 nm, the tyrosine residues are excited principally. When the sample unfolds in 5 M GdnHCl, the fluorescence decreases in intensity because little energy is transferred from the excited tyrosine residues to the sole tryptophan residue: the distances between these residues are greater in the unfolded protein. The poor efficiency of energy transfer in the unfolded state, but not in the folded state, is also evident in the observation that a distinct peak of tyrosine fluorescence is observed for the unfolded protein but not for the folded protein.

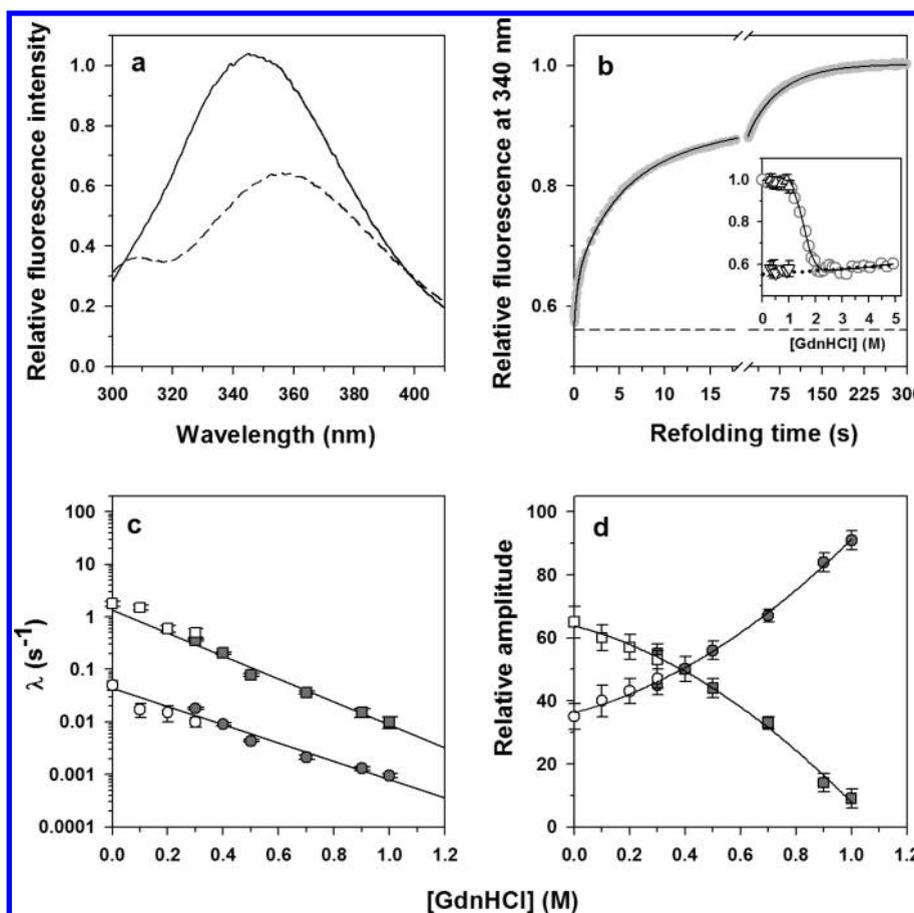
Figure 4b shows the kinetic trace of the refolding of MNEI in 0.3 M GdnHCl, as monitored by fluorescence. The observable change in fluorescence occurs in two exponential kinetic phases. The rate constants of both the kinetic phases as well as their relative amplitudes for folding in 0.3 and 0.7 M GdnHCl are listed in Table 1. The kinetic trace of refolding extrapolates at time zero to the value of the fluorescence signal characteristic of the unfolded protein, indicating that there is no change in the fluorescence signal in the submillisecond time domain. The inset of Figure 4b shows the fluorescence-monitored, GdnHCl-induced equilibrium unfolding curve for MNEI at pH 9.4 and 25 °C. An analysis of the equilibrium unfolding curve according to a two-state  $N \rightleftharpoons U$  model indicates that the midpoint of the equilibrium unfolding transition is 1.5 M GdnHCl, and that the free energy of unfolding is 4.8 kcal mol<sup>-1</sup>. The inset of Figure 4b also compares the kinetic and equilibrium amplitudes of refolding, at different concentrations of GdnHCl, as monitored by fluorescence. The kinetic traces of refolding at different concentrations of GdnHCl extrapolate at time zero to values that fall on the linearly extrapolated unfolded protein baseline of the

equilibrium unfolding transition curve. Hence, the kinetic amplitudes of refolding match the equilibrium amplitudes at all the concentrations of GdnHCl studied.

Figure 4c shows the dependence of the two observed rate constants of refolding on GdnHCl concentration. The logarithm of the observed rate constants of both the kinetic phases, for folding at different concentrations of GdnHCl, increases linearly with a decrease in the concentration of GdnHCl. The observed rate constants in zero denaturant, which were determined by initiating refolding by means of a pH jump from 12 to 9.4, fall on the linearly extrapolated denaturant dependencies of the observed rate constants.

Figure 4d shows how the relative amplitudes of the two kinetic phases vary, when refolding is conducted at different concentrations of GdnHCl. The relative amplitudes of the fast and slow kinetic phases decrease and increase, respectively, with an increase in GdnHCl concentration, at the expense of each other. These results suggest that unfolded MNEI might form native structure on two parallel paths that compete with each other and that the fraction of unfolded molecules utilizing either pathway depends upon the concentration of GdnHCl in which folding commences.

**Cysteine Accessibility-Monitored Refolding Kinetics.** To improve our understanding of the complexities of the folding reaction of MNEI, which is evident from the fluorescence-monitored kinetics, the folding of MNEI was studied using the pulsed SX methodology in conjunction with mass spectrometry. Figure 5 shows the refolding kinetics of MNEI, as monitored by a change in the solvent accessibility of the single thiol group during refolding in 0.3 M GdnHCl (Figure 5a) and 0.7 M GdnHCl (Figure 5b), along with the measured solvent accessibilities in the unfolded and native states. In all cases, the solvent accessibility was measured by determining the extent of labeling by a 4 ms labeling pulse of 2 mM MMTS at pH 9.4. It should be pointed out that this duration of the labeling pulse is more than 500-fold lower than the time constant of the fast folding reaction of MNEI (Table 1). The pulsed SX experiment directly monitors the disappearance of the molecules in which the thiol is exposed to solvent during refolding. It should be noted that the strength and duration of the labeling pulse are just sufficient to label completely the thiol in the unfolded protein (Figure 3).



**Figure 4.** Thermodynamics and kinetics of refolding of MNEI at pH 9.4 and 25 °C as monitored by fluorescence. (a) Fluorescence emission spectra of the protein in 0 M GdnHCl (—) and 5 M GdnHCl (---), with excitation at 280 nm. (b) Kinetic trace of refolding of MNEI in 0.3 M GdnHCl as monitored by the change in fluorescence at 340 nm, after excitation at 280 nm. The black line through the data is a fit to a double-exponential equation, and the black dashed line represents the signal of the protein in 3 M GdnHCl from which refolding was commenced. The inset compares the kinetic and equilibrium amplitudes of refolding. The dark gray empty circles represent data for the equilibrium unfolding transition, and the solid line through the data represents a fit to a two-state  $N \rightleftharpoons U$  model. The triangles represent the  $t = \infty$  signal, and the inverted triangles represent the  $t = 0$  signal, obtained from fitting the kinetic traces of refolding to a double-exponential equation. The black dotted line is a linear extrapolation of the unfolded protein baseline. Panels c and d show the rate constants and relative amplitudes, respectively, of the two phases of refolding, at different concentrations of GdnHCl. In panels c and d, the gray filled squares and gray filled circles represent data for the fast and slow phases of refolding, respectively. The empty squares and empty circles correspond to the two phases of folding kinetics observed for intrinsic tryptophan fluorescence-monitored refolding induced by a pH jump from 12 to 9.4. The solid lines through the data points serve as a guide for the eye only. Error bars, wherever shown, represent the standard deviations from three separate experiments.

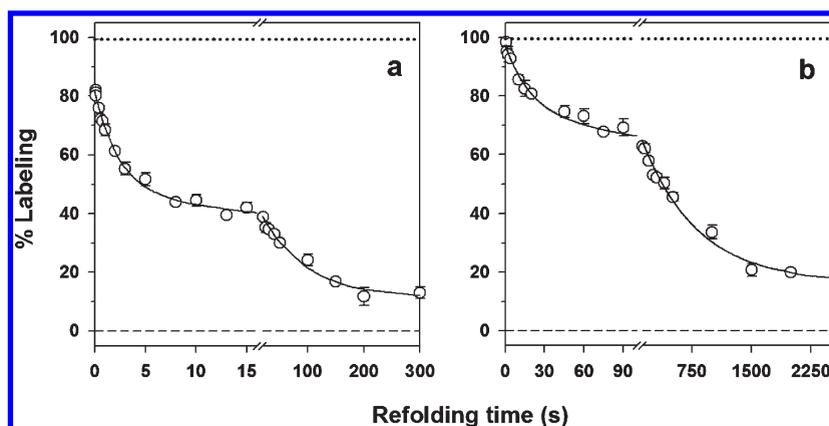
**Table 1. Comparison of the Fluorescence-Monitored and Cysteine Burial-Monitored Kinetics of Refolding of MNEI at pH 9.4 ± 0.1 and 25 °C**

observed phase	0.3 M GdnHCl				0.7 M GdnHCl			
	fluorescence-monitored kinetics <sup>a</sup>		cysteine burial-monitored kinetics		fluorescence-monitored kinetics <sup>a</sup>		cysteine burial-monitored kinetics	
	$\lambda$ (s <sup>-1</sup> )	$\alpha$ (%)	$\lambda$ (s <sup>-1</sup> )	$\alpha$ (%)	$\lambda$ (s <sup>-1</sup> )	$\alpha$ (%)	$\lambda$ (s <sup>-1</sup> )	$\alpha$ (%)
burst phase	—	0	not measured	20 ± 3	—	0	—	0
fast phase	0.36 ± 0.02	55 ± 3	0.4 ± 0.03	35 ± 3	0.04 ± 0.002	33 ± 2	0.05 ± 0.01	22 ± 3
slow phase	0.018 ± 0.002	45 ± 3	0.012 ± 0.001	35 ± 3	0.0021 ± 0.0002	67 ± 2	0.0014 ± 0.0001	58 ± 3
very slow phase	—	0	not measured	10 ± 2	—	0	not measured	20 ± 3

<sup>a</sup> The fast and slow phases account for the entire change in fluorescence that occurs during folding.

Figure 5a shows that the refolding of MNEI in 0.3 M GdnHCl occurs in four kinetic phases, when monitored by measurement of the change in the solvent accessibility of the thiol. Twenty

percent of the protein molecules appear to acquire protection from MMTS labeling very early during refolding (within the burst phase of ~2 ms, ultrafast phase of refolding). This is



**Figure 5.** Kinetics of the change in cysteine accessibility during refolding at pH 9.4 and 25 °C: (a) refolding in 0.3 M GdnHCl and (b) refolding in 0.7 M GdnHCl. Fractional labeling (○) is plotted vs the time of application of the labeling pulse after the commencement of refolding. In each panel, the solid line is a fit of the labeling data to a double-exponential equation. The dotted and dashed lines represent the labeling of unfolded protein in 3 M GdnHCl and native protein in 0.3 M GdnHCl (a) or 0.7 M GdnHCl (b), respectively. Error bars represent the standard deviations of measurements from three different experiments.

followed by observable fast and slow phases in the kinetics of burial of the thiol, in which 70 of the remaining 80% of the molecules acquire protection from MMTS labeling. The observed rate constants of the burial of the thiol, for both the kinetic phases, as well as the relative amplitudes, are listed in Table 1. The observed rate constants for the two kinetic phases as determined from the measurement of the change in solvent accessibility of the thiol and of the change in fluorescence are similar (Table 1). This indicates that both probes report on the same phases of refolding. The relative amplitudes of both the kinetic phases are, however, different when monitored using the two probes (Table 1). Finally, the remaining 10% of the molecules appear to be protected from MMTS labeling in a very slow phase of refolding. The rate of the very slow phase could not be determined precisely, because of its small amplitude.

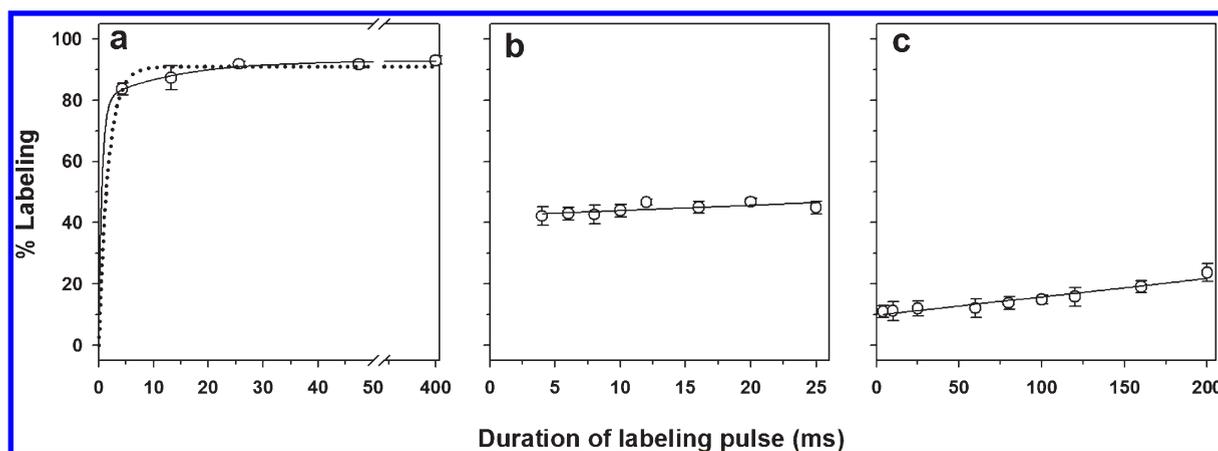
Figure 5b shows that the refolding of MNEI in 0.7 M GdnHCl occurs in three kinetic phases, when monitored by measurement of the change in the solvent accessibility of the thiol. No burst phase protection from MMTS labeling is observed. Twenty percent of the protein molecules appear to acquire protection from MMTS labeling with a rate constant of  $0.05 \pm 0.01 \text{ s}^{-1}$ . Sixty percent of the protein molecules acquire protection from MMTS labeling with a rate constant of  $0.0014 \pm 0.0001 \text{ s}^{-1}$ . The observed rate constants for both the kinetic phases (fast and slow phases) of refolding as determined from the changes in the solvent accessibility of the thiol and from the change in fluorescence are similar, but the relative amplitudes are different (Table 1). The relative amplitudes of the fast and slow phases of refolding in 0.3 M GdnHCl are also different from those of refolding in 0.7 M GdnHCl (Table 1). Finally, the last 20% of the protein molecules appear to be protected from MMTS labeling only in a very slow phase of refolding.

**Determination of the Protection Factor of the Kinetic Intermediate Ensemble.** The observed multiphasic kinetics of burial of the thiol during folding (Figure 5) suggests that either multiple intermediate forms, with differential protection from MMTS labeling, are populated during folding or unfolded MNEI forms native structure via multiple routes. To distinguish between these two possibilities, the protection from chemical labeling of the intermediates populated at the end of the burst, fast and slow phases of folding, was measured. Figure 6a shows

the extent of labeling after refolding for 75 ms in 0.3 M GdnHCl, which corresponds to a time of folding when the burst phase folding reaction is complete and only ~3% of the fast phase folding reaction has occurred. It is seen that the extent of labeling increased from ~80 to ~95% and remained constant thereafter, when the duration of the labeling pulse of 2 mM MMTS was increased from 4–25 to 400 ms. To determine whether the molecules present after folding for 75 ms comprise a single population that is very weakly protected against labeling or of two subpopulations that are unprotected and weakly protected, the data were fit to eqs 11 and 12. The fit of the data to eq 12 appears to be better than that to eq 11, but because of the weak protection (the protection factors obtained from the fits are given in the legend of Figure 6a), it is difficult to be absolutely sure that two subpopulations of unprotected and weakly protected molecules, rather than one population of very weakly protected molecules, are present after folding for 75 ms. Figure 6b shows that the extent of labeling after refolding in 0.3 M GdnHCl for 15 s, which corresponds to ~6 times the time constant of the fast phase, remains constant at ~40–45% even when the duration of the labeling pulse is increased more than 6-fold (from 4 to 25 ms). Figure 6c shows that the extent of labeling after refolding in 0.3 M GdnHCl for 300 s, which corresponds to ~4 times the time constant of the slow phase, increases from only ~10 to ~20%, when the duration of the labeling pulse of 2 mM MMTS is increased from 4 to 200 ms.

## DISCUSSION

**Comparison of the Kinetics of Folding at pH 7 and 9.4.** In this study, the refolding of MNEI was studied at pH 9.4 using the pulsed SX methodology in conjunction with mass spectrometry. It was necessary to conduct the folding studies at pH 9.4 because at this pH it was possible to ensure that all unfolded molecules are labeled by a 4 ms labeling pulse of MMTS (Figure 3). Before examination of the results of the pulsed SX labeling experiments, it is prudent to compare the results of fluorescence-monitored folding experiments at pH 9.4 to those obtained previously at pH 7<sup>30</sup> to determine whether there is a drastic change in the mechanism (Scheme 1) with a change in pH from 7 to 9.4.



**Figure 6.** Dependence of the extent of labeling on the duration of the labeling pulse. Labeling pulses were applied [75 ms (a), 15 s (b), and 300 s (c)] after the initiation of the refolding reaction in 0.3 M GdnHCl. The dotted and solid lines in panel a are nonlinear, least-squares fits of the data to eqs 11 and 12, respectively, which yielded values for  $P_{SX}$  of 3 and 20, respectively.  $k_b$  was set to  $8 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  (see the legend of Figure 3a). The solid lines through the data in panels b and c are drawn by inspection only. The error bars represent the standard deviations of measurements from three different experiments.

The observation that as at pH 7, the relative amplitude of the slow phase increases at the expense of the relative amplitude of the fast phase, with increasing GdnHCl concentration at pH 9.4 (Figure 4), suggests that the two major folding pathways are not affected (Scheme 1). The major difference in the folding reactions at pH 9.4 and 7 is that the change in intrinsic tryptophan fluorescence during folding at pH 9.4 occurs only during the fast and slow kinetic phases of folding (Figure 4). The observation that the very fast phase of fluorescence change is absent at pH 9.4 suggests that the unfolded form  $U_1$  from which the very fast phase originates is populated to a negligible extent at pH 9.4. It implies that the equilibrium distribution of  $U_1$  and  $U_2$  is pH-dependent. It is possible that factors other than the *cis*–*trans* isomerization at X–Pro bonds, such as the ionization of the C42 thiol at high pH, could control the population distribution of  $U_1$  and  $U_2$ . Another possible explanation for the absence of the very fast phase of fluorescence change could be that  $I_{VF}$  (Scheme 1) is greatly destabilized at pH 9.4, while the other folding intermediates are not, but this explanation is unlikely because it would imply that there should be a very slow phase of fluorescence change accompanying the folding of  $U_1$  at pH 9.4, which is not observed.

The very slow phase of folding, leading to the formation of N,<sup>30</sup> continues to remain silent to the intrinsic fluorescence change for folding at pH 9.4. The ultrafast phase of folding has been shown to lead to the formation of a collapsed intermediate ensemble, here called  $I_E$ , even at pH 9.4.<sup>38</sup>

**Burial of C42 Occurs in Four Kinetic Phases.** In the pulsed SX labeling experiments, the strength of the labeling pulse was adjusted so that it was just sufficient to label all unfolded molecules (in which the thiol is solvent-exposed) (Figures 3 and 5), but it is incapable of labeling fully folded molecules (in which the thiol is protected) (Figures 3 and 5). When applied in this manner, the pulsed SX labeling methodology directly reports on the kinetics of the disappearance of the unfolded molecules, or molecules with unfolded-like protection against labeling, during folding. This is a major advantage of utilizing the pulsed SX labeling methodology to study folding.

The kinetics of burial of the C42 thiol during folding in 0.3 M GdnHCl (pH 9.4) displays four kinetic phases (Figure 5). Of

these, the fast and slow phases can also be observed in the kinetics of fluorescence change. Both probes show that the relative amplitude of the slow phase increases at the expense of the fast phase, when folding is conducted at a higher GdnHCl concentration, in agreement with previous studies.<sup>30</sup> The kinetic trace of burial of C42 indicates that  $\sim 20\%$  of the molecules are partially protected in an initial burst phase preceding the fast phase (see below). The mixing dead time (4 ms) is short enough that a kinetic phase of C42 burial corresponding to the very fast phase of fluorescence change would have easily been captured. Hence, the very fast phase of folding cannot be detected at pH 9.4, via measurement of the intrinsic tryptophan fluorescence or via measurement of the C42 thiol reactivity. Thus, it is very likely that the burst phase of C42 burial corresponds to the ultrafast phase of folding detected by ANS fluorescence and small-angle X-ray scattering.<sup>38</sup> At the end of the slow phase,  $\sim 20\%$  of the molecules remain fully capable of being labeled; these molecules appear to be labeled during the very slow phase of folding that had been identified in an earlier study<sup>30</sup> to lead to the formation of N (Scheme 1).

**The Burst Phase of C42 Burial Corresponds to Partial Protection in  $I_E$ .** The intermediate ensemble ( $I_E$ ) appears to be produced by the ultrafast phase of folding of  $U_2$  (Scheme 1). This study shows that after folding in 0.3 M GdnHCl for 75 ms (Figure 5a), when the fast folding reaction has occurred to a negligible extent, only  $\sim 80\%$  of  $I_E$  molecules are labeled by a 4 ms pulse capable of labeling nearly all ( $\sim 98\%$ ) unfolded protein molecules. There could be several explanations. One explanation could be that all  $I_E$  molecules are very weakly protected ( $P_{SX} < 5$ ), and the fraction of molecules that are labeled would then be given by eq 11. Another explanation could be that  $I_E$  is comprised of two subpopulations, with one subpopulation ( $\sim 80\%$ ) having a U-like lack of protection ( $P_{SX} \sim 1$ ) and the second subpopulation ( $\sim 20\%$ ) having weak protection ( $P_{SX} < 20$ ). The fraction of molecules that are labeled would then be given by eq 12. The third explanation could be that  $I_E$  is a strongly protected intermediate ( $P_{SX} \gg 100$ ) that is populated by only 20% of the molecules. In this case, too, the fraction of molecules that would be labeled is given by eq 12, but the strongly protected molecules would be labeled only by a very strong labeling pulse.

While the third explanation appeared to be unlikely because previous studies of folding at pH 9.4 had shown that  $I_E$  contains very little secondary structure and is loosely compact because W4 remains fully solvated within it,<sup>38</sup> it was important to determine the extent of protection in  $I_E$ . Fortunately, it was possible to measure the protection of the thiol in  $I_E$  at a time of refolding (75 ms) when a negligible fraction of  $I_E$  would have undergone the fast folding reaction. The labeling data (Figure 6a) can be explained either on the basis of  $I_E$  being comprised of a single population of molecules that are very weakly protected (with a protection factor of  $\sim 3$ ) or on the basis of  $I_E$  being comprised of two subpopulations of molecules, 80% with no protection against labeling, and the remaining molecules being weakly protected with a protection factor of  $\sim 20$  (see Results). The possibility that a strongly protected intermediate is present after folding for 75 ms is ruled out. The observation that all molecules in  $I_E$  are labeled when folding is conducted in 0.7 M GdnHCl (Figure 5b) indicates that less protective structure forms in  $I_E$  under less stabilizing conditions. ANS binding studies have also shown that when folding is conducted in the presence of  $\geq 0.5$  M GdnHCl, the product of the ultrafast phase has very little solvent-accessible hydrophobic surface capable of binding ANS (A. K. Patra and J. B. Udgaonkar, unpublished results).

**Folding Occurs on Two Parallel Pathways.** The simplest explanation for the observation that for folding in both 0.3 and 0.7 M GdnHCl, the decrease in the extent of labeling of the C42 thiol occurs in two kinetic phases, one fast and another slow, would be that there is one fast phase of folding that leads to the formation of a partially structured intermediate. The fast phase in the C42 burial kinetics would then represent the kinetics of formation of this intermediate, and the slow phase would represent slow labeling of the weakly protected thiol in the partially structured intermediate, with the observable slow rate of labeling being determined by eq 7. It should be noted that if, instead, the intermediate afforded strong protection against C42 labeling, all protein molecules would have been labeled in a single fast phase, and a slow phase of C42 thiol labeling would not have been observed. Two observations suggest that a partially structured intermediate is not formed during the fast phase. (1) The extent of labeling by a pulse applied after folding in 0.3 M GdnHCl for 15 s, when the fast phase is over (see Results), remains at  $\sim 40\%$  when the duration of the pulse is increased from 4 to 25 ms, indicating that the fast phase of folding does not produce a partially structured intermediate that affords only weak protection against labeling at the thiol; instead, a relatively structured intermediate is formed. (2) The apparent rate of the slow phase of C42 burial kinetics matches that of the slow phase of fluorescence change, which is known to accompany a folding reaction.<sup>30</sup> Hence, the observation that there are two kinetic phases, fast and slow, of C42 burial suggests that protein molecules possessing unfolded-like protection disappear via two folding pathways operating in parallel.

It is well-known that fast and slow folding pathways can originate from unfolded forms that differ because of *cis-trans* isomerization at one or more X-Pro bonds.<sup>43,44</sup> Monellin has six Pro residues, of which Pro41 and Pro93 are in the *cis* conformation. In this context, the simplest explanation for the fast and slow folding reactions occurring in parallel would be that they originate from two unfolded forms in equilibrium with each other. The observation made in this study by measurements of fluorescence change as well as of C42 burial, that the amplitude of the slow phase increases at the expense of the fast phase with an

increase in GdnHCl concentration (Figure 4 and Table 1), is, however, inconsistent with the fast and slow folding pathways of monellin originating from two unfolded forms ( $U_2^F$  and  $U_2^S$ ) that are different because of *cis-trans* proline isomerization. This is because such an explanation would demand that the equilibrium between  $U_2^F$  and  $U_2^S$  be fast compared to the fast folding reaction and also be dependent on GdnHCl concentration,<sup>25</sup> but proline isomerization rates and equilibria are known from direct kinetic measurements to be both slow ( $\sim 0.01$  s<sup>-1</sup>)<sup>45-48</sup> and insensitive to GdnHCl concentration.<sup>49-51</sup> Hence, it is unlikely that the fast and slow kinetic phases of folding arise from two unfolded forms that are different in *cis-trans* isomerization of one or more X-Pro bonds. Earlier double-jump interrupted unfolding experiments also had indicated that both the fast and slow folding pathways originate from one unfolded form: both pathways become operational simultaneously when the protein is unfolded transiently,<sup>30</sup> and it is very unlikely that two unfolded states would form simultaneously at the same rate from N, if they differ because of proline isomerization.

The pulsed SX labeling experiments do not, by themselves, suggest the origin of the fast and slow folding pathways. It becomes necessary to examine whether the results of this study are consistent with the results of an earlier study, which are summarized in Scheme 1.<sup>30</sup> An important earlier result<sup>30</sup> was that an early intermediate ensemble,  $I_E$ , forms from one unfolded form (see above) before the commencement of the fast folding reaction. Our study suggests that either  $I_E$  is an ensemble of loosely packed molecules with a protection factor of  $\sim 3$  or it is comprised of two subpopulations, one unprotected (protection factor of  $\sim 1$ ) and one weakly protected with a protection factor of  $< 20$  (see Results). In the earlier study,<sup>30</sup> it had, however, been necessary to postulate that  $I_E$  consists of two subpopulations,  $I_{E1}$  and  $I_{E2}$ , from which the fast and slow folding pathways arise. Our study shows that the products of the fast and slow folding pathways possess structures that afford strong protection against labeling of the C42 thiol, but they also suggest that these products are unlikely to be fully folded protein because the extent of labeling by a pulse applied after refolding for 300 s, increases from only  $\sim 10$  to  $\sim 20\%$ , when its duration is increased from 4 to 200 ms; in contrast, the extent of labeling of native monellin remains zero, for both 4 and 200 ms labeling pulses (Figure 3b, inset). In the earlier study, the fast and slow phases of fluorescence change had been shown to represent the formation of the late intermediates,  $I_F$  and  $I_S$ , respectively (Scheme 1), and because  $I_F$  and  $I_S$  were observed to form in amounts that were not proportional to the rates at which they formed, it had been necessary to postulate that the kinetic partitioning leading to their formation occurred during the formation of  $I_{E1}$  and  $I_{E2}$  from  $U_2$  (Scheme 1). In this study, the observation that the fast and slow phases of C42 burial correspond in rate to the fast and slow phases of fluorescence change, respectively, is consistent with the results of the previous study that indicated that the fast and slow folding reactions produce  $I_F$  and  $I_S$  from  $I_{E1}$  and  $I_{E2}$ , respectively (Scheme 1). The pulsed SX labeling experiments provide structural information about  $I_F$  and  $I_S$ , suggesting that they cannot be only partially structured, thereby affording only weak protection against labeling of the C42 thiol.

The great utility of the pulsed SX experiments is that they provide direct information about the number of molecules folding via each of the two pathways. Only  $\sim 35-50\%$  of the protein molecules have become strongly protected at the end ( $\sim 15$  s) of the fast phase of folding when  $I_F$  would be fully

populated. An additional ~30–45% of the protein molecules become similarly or even more protected only at the end (~300 s) of a slow kinetic phase of folding, when  $I_S$  would be fully formed. At a lower GdnHCl concentration, more molecules are found to be protected at the end of the fast phase than at a higher GdnHCl concentration. Approximately 20% of the protein molecules are unprotected against labeling at 300 s, and it is important to understand why this is so.

To understand the folding of this 20% of protein molecules, it becomes necessary to refer back to Scheme 1 from the previous study.<sup>30</sup> That study had identified a very slow folding phase in which native protein forms from different folding intermediates, including  $I_F$  and  $I_S$ . Taken together, the results of this study and the previous study<sup>30</sup> suggest that the 20% of the protein molecules become protected only during this very slow folding phase. This study shows that the 20% of the molecules that are unprotected against labeling at 300 s of folding cannot be present as U at this time because then the very slow phase of folding would not have been completely silent to fluorescence change, as it is (Figure 4 and Table 1). Instead, they must be present in the form of one or more intermediate forms that already possess the fluorescence of N. The previous study had shown that after folding for 300 s,  $I_F$  and  $I_S$  are the predominant forms present and that both possess nativelike fluorescence.<sup>30</sup> Hence, it appears that a subpopulation of either  $I_F$  or  $I_S$  could be unprotected against labeling. In this study, the observation that the relative amplitudes of the fast and slow phases are different when measured by fluorescence and by C42 burial (Table 1) is suggestive of heterogeneity in  $I_F$ ,  $I_S$ , or both, but it is not possible to conclusively assign the unprotected molecules to either intermediate ensemble at present. Previously, structural heterogeneity had been observed in  $I_F$  but not in  $I_S$ :<sup>30</sup>  $I_F$  was shown to be comprised of two subpopulations,  $I_{F1}$  and  $I_{F2}$  (Scheme 1), both with the fluorescence of N, but with  $I_{F2}$  considerably less stable than  $I_{F1}$ .<sup>30</sup> It is tempting to predict that  $I_{F2}$  is comprised of the population of molecules that are unprotected against labeling, but additional studies are required to establish whether this is really so.

**Importance of Multiple Folding Pathways.** In this study, the use of a novel pulsed SX methodology in conjunction with mass spectrometry, which can directly quantify the changes in the population of molecules with unfolded-like protection, during folding, has shown directly that three intermediates form on parallel folding pathways during the folding of MNEI. The possibility that proteins can use multiple pathways for folding was revealed elegantly in the jigsaw puzzle model for folding,<sup>32</sup> according to which the folding of a protein was likened to the assembly of a jigsaw puzzle. Subsequently, multiple folding pathways have been reported for several proteins, including hen lysozyme,<sup>15</sup> barstar,<sup>16,24</sup> thioredoxin,<sup>23</sup> and a titin domain.<sup>33</sup> It is important to note that for many of these proteins, only one of the several pathways appears to be operative under a given set of experimental conditions. It appears that by making multiple routes available, evolution has ensured the robustness of the folding process.<sup>32</sup> For example, if a mutation or a change in folding conditions disables one folding pathway, another pathway can take over. The result of this study, that the relative utilization of available folding pathways depends upon the solvent conditions, is important because it implies that the folding pathways utilized in the cell depend on the conditions prevalent within it.

Parallel folding pathways arise because of the structural heterogeneity associated with the transition states and

intermediates en route to folding.<sup>54</sup> Transition states may not always be a single population of structures but may consist of subpopulations that differ in their average structure. These subpopulations are likely to fold via independent pathways. Computer simulations suggest that the transition state is not just a single conformation but rather an ensemble of a multitude of conformations<sup>55</sup> and could consist of fluctuating mobile structures.<sup>56</sup> Not surprisingly,  $\phi$ -value analysis of circularly permuted variants of many proteins, including the  $\alpha$ -spectrin SH3 domain<sup>57</sup> and ribosomal protein S6,<sup>58</sup> has shown that the transition state is diffuse and structurally polarized; hence, the proteins fold via different folding pathways. The use of circular permutation in conjunction with  $\phi$ -value analysis has also indicated that activation barriers during the folding and unfolding of proteins can be broad, flat, and malleable and, hence, would appear different under different folding or unfolding conditions.<sup>59–61</sup>

**Conclusion.** In this study, the use of a pulsed SX methodology coupled to mass spectrometry has made it possible to show that unfolded molecules enter two separate folding pathways. The relative utilization of the three pathways depends upon the conditions used for folding. These results elucidate the roles played by folding intermediates in directing the utilization of alternative folding pathways when many pathways are available.

## AUTHOR INFORMATION

### Corresponding Author

\*Telephone: 91-80-23666150. Fax: 91-80-23636662. E-mail: jayant@ncbs.res.in.

### Funding Sources

S.K.J. is a recipient of a Shyama Prasad Mukherjee Fellowship awarded by the Council of Scientific and Industrial Research, Government of India. J.B.U. is a recipient of a J. C. Bose National Research 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.

## ACKNOWLEDGMENT

We thank members of our laboratory for discussion.

## ABBREVIATIONS

U, unfolded; I, intermediate; N, native; pulsed SX, pulsed thiol labeling; MNEI, single-chain monellin; MMTS, methyl methanethiosulfonate; ESI-MS, electrospray ionization mass spectrometry; GdnHCl, guanidinium hydrochloride.

## REFERENCES

- (1) Wetlaufer, D. B. (1973) Nucleation, rapid folding and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* 70, 697–701.
- (2) Kim, P. S., and Baldwin, R. L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51, 459–489.
- (3) Fersht, A. R. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U.S.A.* 92, 10869–10873.
- (4) Nath, U., and Udgaonkar, J. B. (1997) How do proteins fold? *Curr. Sci.* 72, 180–191.
- (5) Daggett, V., and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28, 18–25.

- (6) Poland, D. C., and Scheraga, H. A. (1965) Statistical mechanics of noncovalent bonds in polyamino acids. IX. The two-state theory of protein denaturation. *Biopolymers* 3, 401–419.
- (7) Scheraga, H. A. (1971) Theoretical and experimental studies of conformations of polypeptides. *Chem. Rev.* 71, 195–217.
- (8) Dill, K. A. (1985) Theory for the folding and stability of globular proteins. *Biochemistry* 24, 1501–1509.
- (9) Bryngelson, J. D., and Wolynes, P. G. (1987) Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 84, 7524–7528.
- (10) Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21, 167–195.
- (11) Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4, 10–19.
- (12) Sali, A., Shakhnovich, E., and Karplus, M. (1994) How does a protein fold? *Nature* 369, 248–251.
- (13) Chan, H. S., and Dill, K. A. (1998) Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins: Struct., Funct., Genet.* 30, 2–33.
- (14) Pande, V. S., Grosberg, A., Tanaka, T., and Rokhsar, D. S. (1998) Pathways for protein folding: Is a new view needed? *Curr. Opin. Struct. Biol.* 8, 68–79.
- (15) Radford, S. E., Dobson, C. M., and Evans, P. A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358, 302–307.
- (16) Shastry, M. C., and Udgaonkar, J. B. (1995) The folding mechanism of barstar: Evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* 247, 1013–1027.
- (17) Kiefhaber, T. (1995) Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. U.S.A.* 92, 9029–9033.
- (18) Wildegger, G., and Kiefhaber, T. (1997) Three-state model for lysozyme folding: Triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* 270, 294–304.
- (19) Englander, S. W., Sosnick, T. R., Mayne, L. C., Shtilerman, M., Qi, P. X., and Bai, Y. (1998) Fast and slow folding in cytochrome c. *Acc. Chem. Res.* 31, 737–744.
- (20) Bilsel, O., Zitzewitz, J. A., Bowers, K. E., and Matthews, C. R. (1999) Folding mechanism of the  $\alpha$ -subunit of tryptophan synthase, an  $\alpha/\beta$  barrel protein: Global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* 38, 1018–1029.
- (21) Bieri, O., Wildegger, G., Bachmann, A., Wagner, C., and Kiefhaber, T. (1999) A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding. *Biochemistry* 38, 12460–12470.
- (22) Englander, S. W. (2000) Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* 29, 213–238.
- (23) Bhutani, N., and Udgaonkar, J. B. (2001) GroEL channels the folding of thioredoxin along one kinetic route. *J. Mol. Biol.* 314, 1167–1179.
- (24) Rami, B. R., and Udgaonkar, J. B. (2001) pH-jump-induced folding and unfolding studies of barstar: Evidence for multiple folding and unfolding pathways. *Biochemistry* 40, 15267–15279.
- (25) Wallace, L. A., and Matthews, C. R. (2002) Sequential versus parallel protein-folding mechanisms: Experimental tests for complex folding reactions. *Biophys. Chem.* 101/102, 113–131.
- (26) Wu, Y., and Matthews, C. R. (2002) Parallel channels and rate-limiting steps in complex protein folding reactions: Prolyl isomerization and the  $\alpha$ -subunit of Trp synthase, a TIM barrel protein. *J. Mol. Biol.* 323, 309–325.
- (27) Gianni, S., Travaglini-Allocatelli, C., Cutruzzola, F., Brunori, M., Shastry, M. C., and Roder, H. (2003) Parallel pathways in cytochrome  $c_{551}$  folding. *J. Mol. Biol.* 330, 1145–1152.
- (28) Kamagata, K., Sawano, Y., Tanokura, M., and Kuwajima, K. (2003) Multiple parallel-pathway folding of proline-free staphylococcal nuclease. *J. Mol. Biol.* 332, 1143–1153.
- (29) Krishna, M. M., Lin, Y., and Englander, S. W. (2004) Protein misfolding: Optional barriers, misfolded intermediates, and pathway heterogeneity. *J. Mol. Biol.* 343, 1095–1109.
- (30) Patra, A. K., and 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.
- (31) Bédard, S., Krishna, M. M., Mayne, L., and Englander, S. W. (2008) Protein folding: Independent unrelated pathways or predetermined pathway with optional errors. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7182–7187.
- (32) Udgaonkar, J. B. (2008) Multiple routes and structural heterogeneity in protein folding. *Annu. Rev. Biophys.* 37, 489–510.
- (33) Ballery, N., Desmadril, M., Minard, P., and Yon, J. M. (1993) Characterization of an intermediate in the folding pathway of phosphoglycerate kinase: Chemical reactivity of genetically introduced cysteinyl residues during the folding process. *Biochemistry* 32, 708–714.
- (34) Ha, J. H., and Loh, S. N. (1998) Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat. Struct. Biol.* 5, 730–737.
- (35) Ramachandran, S., Rami, B. R., and Udgaonkar, J. B. (2000) Measurements of cysteine reactivity during protein unfolding suggest the presence of competing pathways. *J. Mol. Biol.* 297, 733–745.
- (36) Sridevi, K., and Udgaonkar, J. B. (2002) Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry* 41, 1568–1578.
- (37) Jha, S. K., and Udgaonkar, J. B. (2007) Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J. Biol. Chem.* 282, 37479–37491.
- (38) Kimura, T., Akiyama, S., Uzawa, T., Ishimori, K., Morishima, I., Fujisawa, T., and Takahashi, S. (2005) Specific collapse followed by slow hydrogen-bond formation of  $\beta$ -sheet in the folding of single-chain monellin. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2748–2753.
- (39) Jha, S. K., Dhar, D., Krishnamoorthy, G., and Udgaonkar, J. B. (2009) Continuous dissolution of structure during the unfolding of a small protein. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11113–11118.
- (40) Jha, S. K., and Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12289–12294.
- (41) Roberts, D. D., Lewis, S. D., Ballou, D. P., Olson, S. T., and Shafer, J. A. (1986) Reactivity of small thiolate anions and cysteine-25 in papain toward methyl methanethiosulfonate. *Biochemistry* 25, 5595–5601.
- (42) Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins. *Adv. Protein Chem.* 21, 287–386.
- (43) Kiefhaber, T., Kohler, H. H., and Schmid, F. X. (1992) Kinetic coupling between protein folding and prolyl isomerization. I. Theoretical models. *J. Mol. Biol.* 224, 217–229.
- (44) Wedemeyer, W. J., Welker, E., and Scheraga, H. A. (2002) Proline cis-trans isomerization and protein folding. *Biochemistry* 41, 14637–14644.
- (45) Grathwohl, C., and Wüthrich, K. (1981) NMR studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* 20, 2623–2633.
- (46) Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J. Mol. Biol.* 279, 449–460.
- (47) Bosco, D. A., Eisenmesser, E. Z., Pochapsky, S., Sundquist, W. I., and Kern, D. (2002) Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5247–5252.
- (48) Shi, T., Spain, S. M., and Rabenstein, D. L. (2004) Unexpectedly fast cis/trans isomerization of Xaa-Pro peptide bonds in disulfide-constrained cyclic peptides. *J. Am. Chem. Soc.* 126, 790–796.
- (49) Nall, B. T., Garell, J.-R., and Baldwin, R. L. (1978) Test of the extended two-state model for the kinetic intermediates observed in the folding transition of ribonuclease A. *J. Mol. Biol.* 118, 317–330.
- (50) Schmid, F. X., and Baldwin, R. L. (1979) The rate of interconversion between the two unfolded forms of ribonuclease A does not depend on guanidinium chloride concentration. *J. Mol. Biol.* 133, 285–287.
- (51) Sridevi, K., and Udgaonkar, J. B. (2003) Surface expansion is independent of and occurs faster than core solvation during the unfolding of barstar. *Biochemistry* 42, 1551–1563.

(52) Harrison, S. C., and Durbin, R. (1985) Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. U.S.A.* 82, 4028–4030.

(53) Wright, C. F., Lindorff-Larsen, K., Randles, L. G., and Clarke, J. (2003) Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.* 10, 658–662.

(54) Jha, S. K., and Udgaonkar, J. B. (2010) Free energy barriers in protein folding and unfolding reactions. *Curr. Sci.* 99, 457–475.

(55) Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. (1996) Protein folding funnels: The nature of the transition state ensemble. *Folding Des.* 1, 441–450.

(56) Guo, Z., and Thirumalai, D. (1997) The nucleation-collapse mechanism in protein folding: Evidence for the non-uniqueness of the folding nucleus. *Folding Des.* 2, 377–391.

(57)iguera, A. R., Blanco, F. J., and Serrano, L. (1995) The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. *J. Mol. Biol.* 247, 670–681.

(58) Lindberg, M. O., Tangrot, J., and Oliveberg, M. (2002) Complete change of the protein folding transition state upon circular permutation. *Nat. Struct. Biol.* 9, 818–822.

(59) Silow, M., and Oliveberg, M. (1997) High-energy channeling in protein folding. *Biochemistry* 36, 7633–7637.

(60) Olofsson, M., Hansson, S., Hedberg, L., Logan, D. T., and Oliveberg, M. (2007) Folding of S6 structures with divergent amino acid composition: Pathway flexibility within partly overlapping foldons. *J. Mol. Biol.* 365, 237–248.

(61) Lindberg, M. O., and Oliveberg, M. (2007) Malleability of protein folding pathways: A simple reason for complex behavior. *Curr. Opin. Struct. Biol.* 17, 21–29.

(62) DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.