BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

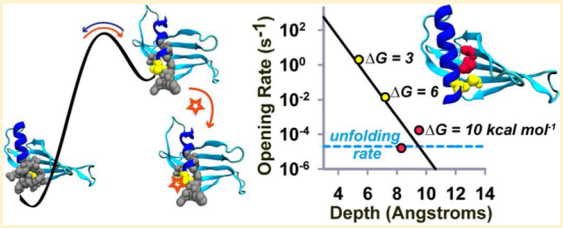
High-Energy Intermediates in Protein Unfolding Characterized by Thiol Labeling under Nativelike Conditions

Pooja Malhotra and Jayant B. Udgaonkar*

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

Supporting Information

ABSTRACT: A protein unfolding reaction usually appears to be so dominated by a large free energy barrier that identifying and characterizing high-energy intermediates and, hence, dissecting the unfolding reaction into multiple structural transitions have proven to be a challenge. In particular, it has been difficult to identify any detected high-energy intermediate with the dry (DMG) and wet (WMG) molten globules that have been implicated in the unfolding reactions of at least some proteins. In this study, a native-state thiol labeling methodology was used to identify high-energy intermediates, as well as to delineate the barriers to the disruption of side chain packing interactions and to site-specific solvent exposure in different regions of the small protein, single-chain monellin (MNEI). Labeling studies of four single-cysteine-containing variants of MNEI have identified three high-energy intermediates, populated to very low extents under nativelike conditions. A significant dispersion in the opening rates of the cysteine side chains has allowed multiple steps, leading to the loss of side chain packing, to be resolved temporally. A detailed structural analysis of the positions of the four cysteine residue positions, which are buried to different depths within the protein, has suggested a direct correlation with the structure of a DMG, detected in previous studies. It is observed that side chain packing within the core of the protein is maintained, while that at the surface is disrupted, in the DMG. The core of the protein becomes solvent-exposed only in a WMG populated after the rate-limiting step of unfolding at high denaturant concentrations.



The unfolding of at least some proteins[1−7] appears to commence by expansion of the native protein (N) structure leading to the transient formation of a dry molten globule (DMG) intermediate. In the DMG, some tight packing interactions have broken,[8−10] but the protein core has not yet become hydrated.[5] The principal loss of structure occurs as the DMG further unfolds with the penetration of water, and although this may occur by gradual diffusive unfolding of the protein,[11] there is little structural information available for any protein about the sequence of unfolding events in what appears to be the rate-limiting step of unfolding to the unfolded state (U).[1−3,5] An important unanswered question concerning the unfolding of a DMG is whether water penetrates into its structure in an all-or-none manner or whether it does so gradually. When does the core become fully hydrated? Are packing interactions lost sequentially or cooperatively? Such questions are difficult to answer because of the difficulty in even identifying the presence of high-energy unfolding intermediates. Characterizing their structures is an even greater challenge.

Intermediates higher in energy than the N, DMG, and U states undoubtedly exist, but they are too transiently populated to be studied by conventional spectroscopic methods. Although they can be stabilized by tuning either the protein sequence by mutation[12−14] or the experimental conditions,[15,16] structural characterization at high resolution is nevertheless difficult. Intermediates that are populated at low levels, ≥0.5%, can now, however, be characterized on the millisecond time scale by newer nuclear magnetic resonance (NMR) methods.[17] The native-state hydrogen exchange (HX) method[18−22] can also be used to detect high-energy intermediates, but it provides structural information only at the level of the protein main chain and not side chains. The analogous native-state thiol labeling (SX)[23,24] method allows characterization of the intermediate structure at the level of side chain packing and solvent accessibility. SX methods in general have proven to be useful in providing site-specific kinetic and thermodynamic information about the exposure of cysteine side chains engineered into different parts of the protein, during either protein folding[25−28] or unfolding.[29,30] The capability of SX experiments, unlike HX experiments,[21] to determine equilibrium and kinetic data under exactly the same conditions makes it possible to place structure opening and closing events in the context of global unfolding and folding events. The native-state SX method appears to be well suited to the characterization of the progressive decrease in the level of side chain packing and increase in side chain solvent accessibility that occurs when a DMG intermediate unfolds to the U state.

One of the proteins that has been shown to commence unfolding by transforming to a DMG is the small protein, single-chain monellin (MNEI), which is derived[31] from its naturally occurring double-chain variant (dcMN).[32] In

Received: November 5, 2013
Revised: April 7, 2014
Published: May 15, 2014

monellin, a single α-helix is packed against five antiparallel β-strands in a β-grasp fold.[33] The unfolding and refolding reactions of MNEI have been characterized extensively.[3,11,28,34−37] In the DMG of MNEI, the α-helix has moved marginally away from the β-sheet, but little is known about further unfolding, other than it can occur as a gradual diffusive process[11] on more than one pathway.[11,35,37] The native-state SX methodology can potentially provide structural, kinetic, and thermodynamic information about the movement of the α-helix away from the β-sheet in the DMG and upon further unfolding.

In this study, the native-state SX methodology has been used on four single-cysteine-containing variants of MNEI, in each of which the cysteine side chain is buried at a different location at the helix−β-sheet interface (Figure 1A). Three high-energy

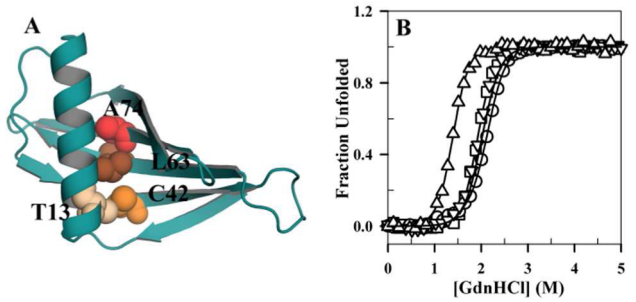


Figure 1. Characterization of the single-cysteine-containing mutant forms of monellin. (A) Structure of monellin, showing residues A74, L63, and T13, which were individually mutated to cysteine, as well as C42, which is present in the wild-type protein. The structure was drawn from PDB entry 1IV7 using PyMOL. (B) Equilibrium unfolding transitions of Cys74 (○), Cys63 (△), Cys42 (□), and Cys13 (▽) at pH 8.0 and 25 °C, monitored by fluorescence emission at 340 nm, upon excitation at 280 nm. The solid lines through the data represent nonlinear least-squares fits to a two-state unfolding model.

intermediates, populated to very low extents under nativelike conditions, have been identified. The kinetic and thermodynamic data obtained have allowed a mapping of these partially unfolded forms on both sides of the main rate-limiting barrier, under different denaturant conditions. The rates at which cysteine residues, located at different locations along the helix−sheet interface, become accessible to labeling and hence to solvent are dispersed over a 10^5-fold range, indicating that solvent accessibility increases in multiple steps, and have allowed the structural events accompanying the unfolding reaction to be resolved temporally. The cysteine residue in the α-helix becomes exposed presumably because of the fluctuating motion of the helix away from the β-sheet, faster than residues in the β-sheet. Side chains in the core remain protected until after the rate-limiting step. The characterization of the kinetics and/or thermodynamics of site-specific labeling of cysteine side chains under nativelike conditions has allowed a delineation of when and how packing interactions are lost during the unfolding of the DMG, to allow penetration of the solvent into the core of the protein.

MATERIALS AND METHODS

Protein Purifications. The method for the purification of MNEI has been described previously.[35] The wild-type protein (Cys42) has a single exposed tryptophan (W4) residue in the first β-strand, a single buried cysteine (C42) in the second β-strand, and seven tyrosine residues in different parts of the structure. The single-cysteine-containing mutant proteins, C42AA74C (Cys74), C42AL63C (Cys63), and C42AT13C (Cys13), were generated by site-directed mutagenesis and purified using the same protocol that was used for the wild-type protein. The purity of each protein was confirmed by mass spectrometry. The concentrations were determined by measuring the absorbance at 280 nm, using an extinction coefficient of 14600 M^{-1} cm^{-1}.[35]

Reagents. All experiments were conducted at pH 8.0 and 25 °C. The reagents used in the experiments were of the highest purity grade from Sigma. Guanidine hydrochloride (GdnHCl) was purchased from USB and was of the highest purity grade. Native buffer contained 100 mM Tris and 500 μM EDTA, and unfolding buffer contained, in addition, GdnHCl. Concentrations of GdnHCl solutions were determined by measurement of the refractive index on an Abbe 3L refractometer from Milton Roy. Concentrations of the stock solutions of PDS and DTNB were measured using their molar absorption coefficients at 247 nm (ε_{247} = 16300 M^{-1} cm^{-1}) and 325 nm (ε_{325} = 17380 M^{-1} cm^{-1}), respectively.[38−40] All buffers and solutions were filtered through 0.22 μm filters and degassed before being used.

Fluorescence-Monitored Equilibrium and Kinetic Studies. GdnHCl-induced equilibrium unfolding transitions were monitored on the stopped-flow module (SFM 4) from Biologic. An excitation wavelength of 280 nm was used, and emission was collected at 340 nm using a 10 nm band-pass filter (Asahi Spectra). The final protein concentration was 10 μM. A detailed description is given in the Methods of the Supporting Information.

Measurement of Thiol Labeling. 4,4′-Dipyridyl disulfide (PDS) and 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) were used as the thiol labeling reagents. They react with a solvent-exposed cysteine residue in the protein to produce a mixed disulfide and a free moiety (4-thiopyridone and TNB^-, respectively). The labeling kinetics was measured by following the increase in absorbance due to the released moiety, at 324 nm for PDS and 412 nm for DTNB labeling.[38−40] The slow labeling reactions were monitored by manual mixing, with a dead time of 20 s, on a CARY 300 double-beam spectrophotometer; the faster reactions were monitored on the SFM 4 module, with a dead time of 10 ms. Labeling of the protein, which was equilibrated in different concentrations of GdnHCl for 6 h, was initiated by addition of the labeling reagent at the same denaturant concentration, so that the final concentration of GdnHCl remained constant.[23] Because thiol labeling reactions are sensitive to small changes in pH, the pH was carefully checked and found not to have changed at the end of the labeling reaction. It should also be mentioned here that the addition of PDS to a buffer solution containing GdnHCl did not cause any appreciable changes in pH. The final protein concentration was 5−15 μM for PDS labeling and 30 μM for DTNB labeling. The concentration of the labeling reagent was at least 5-fold higher than the protein concentration, in all experiments, to maintain pseudo-first-order conditions. To rule out any artifact that might arise from the labeling reagent possibly sticking to the sides of the reaction tube, the concentration of the reagent was checked in the blank solution at the end of the reaction and was not found to be significantly altered. Labeling kinetics was also measured at higher GdnHCl concentrations (above the midpoint of denaturant-induced unfolding, C_m), in the case of Cys13, by diluting native protein

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

into a final high concentration of denaturant (2−7 M GdnHCl), in the presence of an at least 10-fold excess of PDS.

It should be noted that for both Cys74 and Cys63, precipitation was observed after 10 h. Because the labeling kinetic traces were found to reach the expected amplitude, indicating that complete labeling had occurred, it was concluded that it was the labeled protein that slowly precipitated. This is also consistent with the observation that at higher PDS concentrations, at which the labeling is faster, the precipitation was also found to be faster. However, the precipitation was slow enough not to affect the measurement of the labeling kinetics.

Determination of the Bimolecular Rate Constant of Labeling. The labeling reaction of an exposed cysteine residue was monitored to determine the bimolecular rate constant (k_b). Because the cysteine in the protein is exposed only at GdnHCl concentrations exceeding the C_m, k_b was measured at lower denaturant concentrations with pentapeptides having the same flanking sequence present in the protein (FRCDI for Cys74, YQCYV for Cys63, RPCMK for Cys42, and PFCQN for Cys13), to account for the effect of local chemistry on the labeling reaction.[23] The peptides were purchased from Genscript Corp. and further purified to ≤98% purity by reverse phase chromatography. The labeling kinetics of the peptides was measured under conditions similar to those used for the protein.

Data Analysis. *Analysis of Equilibrium and Kinetic Data.* The equilibrium unfolding transitions were analyzed according to a two-state N ↔ U model.[41] The free energy of unfolding, ΔG_u, is assumed to be linearly dependent on denaturant concentration:

$$\Delta G_u = \Delta G_u^{H_2O} - m_G[\text{denaturant}] \quad (1)$$

where m_G is the denaturant dependence of ΔG_u and $\Delta G_u^{H_2O}$ is the free energy of unfolding in water.

For a two-state process, the logarithm of the rate constants of unfolding and refolding shows a linear dependence on denaturant concentration, as described by the equations

$$\ln \lambda_u = \ln \lambda_u^{H_2O} + m_{\lambda_u}[\text{denaturant}] \quad (2)$$

$$\ln \lambda_f = \ln \lambda_f^{H_2O} + m_{\lambda_f}[\text{denaturant}] \quad (3)$$

where $\lambda_u^{H_2O}$ and $\lambda_f^{H_2O}$ are the unfolding and refolding rates in 0 M GdnHCl, respectively, and m_{λ_u} and m_{λ_f} represent their dependences on denaturant concentration, respectively.

Analysis of Thiol Labeling Kinetics. Under pseudo-first-order conditions, the observed rate of labeling, k_{obs}, of an exposed cysteine in the unfolded state, or an unstructured peptide, by a thiol reagent L, is given by $k_{obs} = k_b[L]$, where k_b is the bimolecular rate constant of labeling. The value of k_b is therefore obtained as the slope of a plot of k_{obs} versus [L].

The labeling of a cysteine residue, buried in the native state, can be described according to the thiol labeling (SX) mechanism,[23,25] similar to the Linderstrøm−Lang mechanism for amide hydrogen exchange reactions,[42] as shown in Scheme 1. The labeling of a buried cysteine in the closed state is limited by the opening rate (k_{op}) and the closing rate (k_{cl}). The exchange competent form [Open(-S-H)] can be formed due to local, subglobal, or global unfolding events that expose the cysteine side chain. Under steady-state conditions, the observed rate of labeling, k_{obs}, is given by

$$k_{obs} = \frac{k_{op}k_b[L]}{k_{op} + k_{cl} + k_b[L]} \quad (4)$$

Under nondenaturing conditions, which do not favor opening events, $k_{op} \ll k_{cl}$. Equation 4 simplifies under two limiting conditions, depending on the relative values of k_{cl} and $k_b[L]$.

In the SX1 limit, when $k_{cl} \ll k_b[L]$, then

$$k_{obs} = k_{op} \quad (5)$$

Hence, in the SX1 limit, k_{obs} is independent of [L], and k_{op} is directly measured.

In the SX2 limit, when $k_{cl} \gg k_b[L]$, then

$$k_{obs} = \frac{k_{op}k_b[L]}{k_{cl}} = K_{op}k_b[L] \quad (6)$$

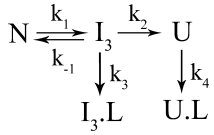
Hence, in the SX2 limit, k_{obs} shows a linear dependence on [L]. The equilibrium constant for opening ($K_{op} = k_{op}/k_{cl}$) and, thus, the free energy associated with opening [$\Delta G_{op} = -RT \ln(K_{op})$] are measured. Hence, by varying [L], kinetic as well as thermodynamic information can be obtained.

Kinetic Simulations. Scheme 2 was used to simulate the fluorescence and thiol labeling data for Cys74, using KINSIM.[43]

Scheme 1

$$\text{Closed (-S -H)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{Open (-S-H)} \xrightarrow[[L]]{k_b} \text{Exchanged (-S-S-X)}$$

Scheme 2



According to Scheme 2, N forms an on-pathway, labeling competent form I_3, which finally converts into U. The rate constants k_1 and k_{-1} correspond to k_{op} and k_{cl}, respectively, while k_3 and k_4 represent the rate of labeling ($k_{obs} = k_b[L]$) of the exposed cysteine in I_3 and U. All the rates except k_2, which is the rate of conversion from I_3 to U, were fixed at the experimentally measured values in the simulations.

RESULTS

Characterization of the Mutant Proteins. Figure 1B compares the equilibrium unfolding curves of the four single-cysteine-containing mutant forms of monellin (Cys74, Cys63, Cys42, and Cys13) used in this study. The free energies of unfolding obtained from these data are compared in Table 1. While Cys13, Cys42, and Cys74 are similar in stability, Cys63 is found to be destabilized by 1.8 kcal mol^{-1}. It appears that the replacement of leucine at position 63 with cysteine may have created a destabilizing cavity within the protein.

Thiol Labeling Kinetics. The labeling of the C42 thiol in Cys42 by PDS was studied under pseudo-first-order conditions (Figure 2). In the absence of any denaturant, the labeling of the thiol in the native protein is described well by a single exponential. In 3 M GdnHCl, where the protein is fully unfolded, the labeling of the thiol in the unfolded protein occurs again with single-exponential kinetics, at a rate much

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

Table 1. Comparison of the Kinetic and Thermodynamic Parameters Measured by Fluorescence and Thiol Labeling at pH 8.0 and 25 °C[a]

	fluorescence			thiol labeling		
protein	$\Delta G_u^{H_2O}$ (kcal mol^{-1})	$\lambda_{vf}^{H_2O}$[b] (s^{-1})	$\lambda_u^{H_2O}$ (s^{-1})	$\Delta G_{op}^{H_2O}$ (kcal mol^{-1})	$k_{cl}^{H_2O}$ (s^{-1})	$k_{op}^{H_2O}$ (s^{-1})
Cys74	6.5 ± 0.19 (3.1)	436 (−4.8)	1.6×10^{-5} (0.7)	10.6 (−3.1)	1100 (−2.3)	1.58×10^{-5} (0.7)
Cys63	4.4 ± 0.21 (3.3)	148 (−5.7)	4×10^{-4} (0.7)	8.3 ± 0.15 (−3.2)	183 (−2.2)	1.7×10^{-4} (0.6)
Cys42	6.2 ± 0.18 (3.2)	115 (−4.2)	5×10^{-5} (0.7)	6.9 ± 0.1 (−0.3)	1790	1.4×10^{-2}
Cys13	6 ± 0.25 (3)	–	3.2×10^{-5} (0.7)	2.7 ± 0.1 (−0.4)	>2000	>2

[a]Values in parentheses are the corresponding *m* values in kilocalories per mole per molar. [b]vf stands for the very fast phase of refolding observed by fluorescence.

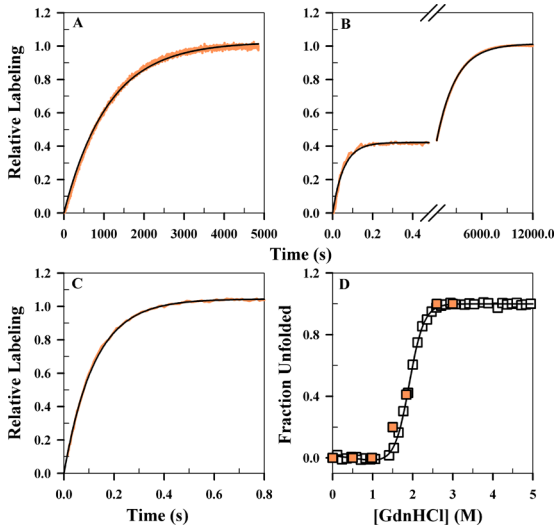


Figure 2. Representative traces of thiol labeling of Cys42 by PDS at pH 8.0 and 25 °C. The native protein was equilibrated at different GdnHCl concentrations, and labeling was initiated by the addition of PDS, without any change in the denaturant concentration. The orange lines represent the labeling traces and the solid black lines the fits. (A) Labeling of the native protein in 0 M GdnHCl by 3 mM PDS. The solid black line shows a single-exponential fit of the data; only a slow phase of labeling is seen. (B) Labeling of the protein equilibrated in 1.84 M GdnHCl by 0.7 mM PDS. The solid black line is a fit to the sum of two exponentials; a slow phase and a fast phase of labeling are seen. (C) Labeling of the protein incubated in 3 M GdnHCl by 0.5 mM PDS. The solid black line is a single-exponential fit of the data; only a fast phase of labeling is seen. (D) Relative amplitude of the fast phase of labeling. The relative amplitude of the fast phase of labeling (red squares) is plotted together with the fraction of unfolded protein, f_U (□), at different GdnHCl concentrations. f_U values were determined from the equilibrium unfolding curve shown in Figure 1.

faster than that observed for the native protein. In 1.84 M GdnHCl, in which approximately half the Cys42 molecules are unfolded at equilibrium, labeling occurs with biexponential kinetics, with the slow apparent rate constant similar to that observed for the native protein, and the fast rate constant similar to that observed for the unfolded protein. The final absorbance signal indicates the extent of labeling. More than 90% labeling was observed under all experimental conditions.

The relative amplitudes of the fast phase of labeling, corresponding to the labeling of the thiol in protein in the U state or in a state in which the thiol has U-like exposure, and of the slow phase corresponding to the labeling of the thiol in protein in the N state were determined at the different GdnHCl concentrations defining the transition zone of the equilibrium unfolding curve. The relative amplitude of the fast phase of labeling was found to predict the fraction of unfolded protein, determined from the equilibrium unfolding curve (Figure 2D). Hence, it appears that the fast phase corresponds to labeling of the cysteine thiol in U, where it would be completely solvent-exposed. The relative amplitudes of the slow and fast phases did not change when the concentration of PDS was varied in the micromolar to millimolar range, indicating that the presence of the thiol reagent does not affect the stability of the protein.

The values of the bimolecular labeling rate constant, k_b, of the cysteine thiol in the protein in the U state were determined for different GdnHCl concentrations defining the unfolded protein baseline of the equilibrium unfolding curve (see Materials and Methods). Figure S1 of the Supporting Information shows the values of k_b for unfolded Cys74, Cys63, Cys42, and Cys13. It also shows values of k_b determined at lower GdnHCl concentrations using a pentapeptide for each protein, whose sequence flanking the cysteine on both sides corresponded to the sequence flanking the cysteine residue in the protein. It is seen that for each protein, the values obtained using the peptide at low GdnHCl concentrations merge with the values obtained using the unfolded protein at high denaturant concentrations. Hence, the values of k_b, obtained using the pentapeptide at low GdnHCl concentrations, are applicable for a fully exposed thiol in the protein at low GdnHCl concentrations. This was important because proper application of eq 4 to describe the labeling of the thiol in the native protein requires a reliable value of k_b at each of the GdnHCl concentrations in which native-state SX measurements were conducted. Figure S1 of the Supporting Information also shows that the values of k_b are similar for the cysteine thiols in the different locations in the four different proteins. In each case, the value of k_b decreases with an increasing GdnHCl concentration, presumably because of the accompanying increase in viscosity.[23]

The kinetics of labeling of the cysteine side chain in the native protein under nativelike conditions (zero to low GdnHCl concentrations) has been analyzed using eq 4. In the analysis, the value of k_b was fixed at the value measured for the exposed cysteine thiol, in either U or an unstructured pentapeptide. In native-state SX experiments, it is assumed that the cysteine thiol is as exposed in the labeling competent form as it is in U.[23,44] In native-state HX experiments, too, it is assumed that the intrinsic exchange rate is the same as that measured for unstructured model peptides.[45]

Native-State SX Measurements. The observed rate of labeling of the cysteine thiol in native protein, k_{obs}, is expected to vary linearly with PDS concentration in the SX2 limit (eq 6) and to be independent of PDS concentration in the SX1 limit (eq 5). Figure 3 shows how k_{obs} varies with PDS concentration, for each of the single-cysteine-containing proteins at different GdnHCl concentrations. Depending on the values of $k_B[L]$, k_{cl}, and k_{op} for a given single-cysteine-containing protein, and

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

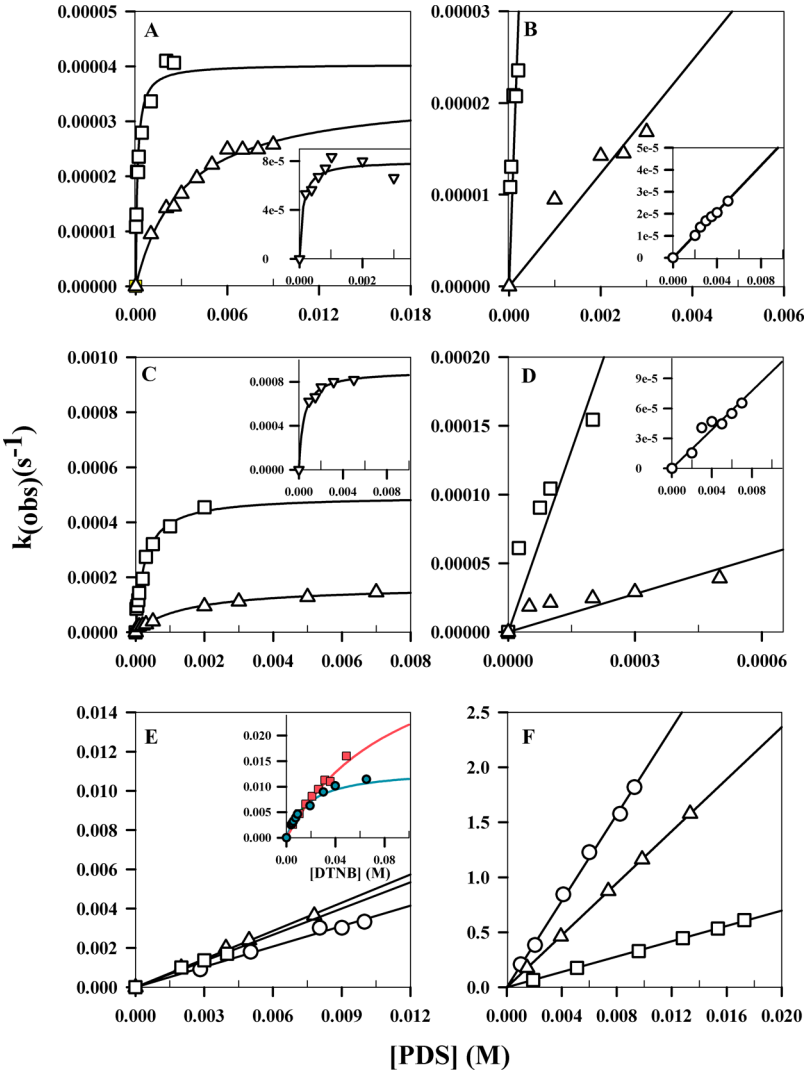


Figure 3. Thiol labeling kinetics at pH 8.0 and 25 °C. The observed rates of thiol labeling (k_{obs}), obtained by fitting the traces corresponding to the slow labeling phase (i.e., the labeling of native protein) to a single-exponential equation, are shown as a function of PDS concentration. (A and B) Cys74 in 0.75 (△) and 1 (□) M GdnHCl. The insets show k_{obs} values in 1.25 (▽) and 0.23 (○) M GdnHCl, respectively. (C and D) Cys63 in 0.25 (△) and 0.5 (□) M GdnHCl. The insets show k_{obs} values in 1 (▽) and 0 (○) M GdnHCl, respectively. (E) Cys42 in 0 (○), 1 (△), and 1.5 (□) M GdnHCl. (F) Cys13 in 0 (○), 0.6 (△), and 1.84 (□) M GdnHCl. The inset in panel E shows the observed rates of labeling for Cys42 with DTNB in 0 (blue circles) and 0.2 (red squares) M GdnHCl. The solid lines through the data in panels A and C represent nonlinear least-squares fits to eq 4, and the solid lines through the data in panels B and D−F represent linear regression fits to eq 6. For all fits, the values of k_b were constrained to the measured values (Figures S1 and S2 of the Supporting Information).

because these rates depend on the concentration of PDS and GdnHCl used, labeling was observed to happen in both the SX1 and SX2 limits, in only the SX2 limit, or primarily in the SX1 limit.

Figure 3A shows the dependence of k_{obs} on PDS concentration for the labeling of Cys74 in 0.75 and 1 M GdnHCl, and Figure 3C does likewise for the labeling of Cys63 in 0.25 and 0.5 M GdnHCl. In these cases, the values of k_{obs} that could be determined cover 70−80% of the total observed range of k_{obs} values. Hence, these data could be fit reliably to eq 4, and dependable values of k_{cl} and k_{op} and hence for K_{op} and ΔG_{op} could be obtained. To confirm that the values obtained for K_{op} at low (<1 M for Cys74 and <0.5 M for Cys63) GdnHCl concentrations are dependable, the values obtained for k_{obs} at the lowest PDS concentrations were fit to eq 6 for the SX2 limit (Figure 3B,D). It was found that these data for both proteins are described well by eq 6, and from the slopes of the straight lines, the values of K_{op} (and ΔG_{op}) could be determined. For both proteins, the values obtained for K_{op} from analyzing k_{obs} values only in the SX2 limit agreed very well with the values obtained by fitting the entire dependence of k_{obs} on PDS concentration. For the labeling of Cys74 in 0.25 M GdnHCl, as well as for the labeling of Cys63 in 0 M GdnHCl, the values obtained for k_{obs} are all seen to be in the SX2 limit (insets of panels B and D of Figure 3), and dependable values of K_{op} could be obtained. It should be noted that it was not possible to use PDS concentrations of >10 mM, because of solubility constraints, and hence, the SX1 limit of labeling could not be reached in 0.25 M GdnHCl for Cys74 or 0 M GdnHCl for Cys63.

At high concentrations (>1.25 M for Cys74 and >1 M for Cys63) of GdnHCl, however, the values determined for k_{obs} covered only the last 30% of the total range of k_{obs} values (insets of panels A and C of Figure 3), and although a reliable

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

value of k_{op} is obtained at high PDS concentrations where labeling occurs primarily in the SX1 limit, it was possible that the values obtained for k_{cl} and, hence, for K_{op} by fitting the data to eq 4 may not be as dependable. When, however, only the last 30% of the dependencies of k_{obs} on PDS concentration, in 0.75 and 1 M GdnHCl for Cys74 and in 0.25 and 0.5 M GdnHCl for Cys63, were analyzed according to eq 4, the values obtained for K_{op} were found to be similar to the values obtained by fitting the complete dependencies of k_{obs} on PDS concentration and to be similar to the values obtained by fitting only that data that were in the SX2 limit (see above and below). This suggests that the values of K_{op} obtained at high GdnHCl concentrations for both proteins are unlikely to be unreliable.

In the case of Cys42, labeling in 0 M GdnHCl was found to take place in the SX2 limit for PDS concentrations of up to 10 mM (Figure 3E), above which the solubility of PDS limited the measurements. Labeling in low concentrations of GdnHCl of up to 2 M was also found to be predominantly in the SX2 limit; at higher concentrations of PDS, the dependence of k_{obs} on PDS concentration decreases, indicating that labeling was now occurring by both the SX2 and SX1 mechanisms. Consequently, only data for lower concentrations of PDS (Figure 3E), where the dependence of k_{obs} on PDS concentration is linear, were fit to eq 6 to obtain values of K_{op} and ΔG_{op} at different GdnHCl concentrations (see below).

In an attempt to obtain labeling data extending into the SX1 limit, a second thiol labeling reagent, DTNB, which has a higher solubility, was also used. The inset of Figure 3E shows that at zero denaturant, k_{obs} has begun to saturate at the highest concentration of DTNB (60 mM) that could be used. Hence, the data were fit to eq 4, using the k_b measured for the DTNB labeling of Cys42 (Figure S2 of the Supporting Information). Values obtained for k_{op}, k_{cl}, and ΔG_{op} at zero denaturant are listed in Table 1. Unfortunately, it was not possible to obtain values of k_{op} and k_{cl} from DTNB labeling of Cys42 in the presence of GdnHCl: even in 0.2 M GdnHCl, the labeling kinetics remains in the SX2 limit (Figure 3E, inset). Values of K_{op} could, however, be obtained at different concentrations of GdnHCl of up to 2 M, by fitting the linear part of the dependencies of k_{obs} on DTNB concentration to eq 6. The values obtained for K_{op} from PDS labeling match those obtained from DTNB labeling at all GdnHCl concentrations (see below). Compared to PDS, DTNB is a larger, charged molecule, and thiol labeling is known to depend significantly on the chemistry of the labeling reagent.[23,25,27] Hence, the observation that similar values of ΔG_{op} at zero denaturant (6.8 kcal mol^{-1} for PDS labeling and 6.9 kcal mol^{-1} for DTNB labeling) are obtained using the two reagents reinforces the reliability of the values of the kinetic and thermodynamic parameters obtained using the SX methodology.

In the case of Cys13, k_{obs} increases linearly with an increase in PDS concentration up to 20 mM (Figure 3F), which is the solubility limit for this reagent at higher denaturant concentrations. Thus, for Cys13, labeling takes place entirely in the SX2 limit, and hence, only the values of K_{op} and ΔG_{op} could be determined by fitting the data to eq 6 (Table 1). Nevertheless, because k_{obs} in the SX1 limit (equal to the k_{op}) has to be necessarily higher than any value of k_{obs} measured in the SX2 limit, a lower limit for k_{op} can be set at 2 s^{-1}. Given the value of K_{op}, a lower limit of 2000 s^{-1} can then be set for k_{cl} (Table 1).

Global Folding and Unfolding Kinetics. It was important to place the site-specific cysteine thiol opening and closing rates in the context of the global folding and unfolding rates measured by fluorescence. In previous studies,[35] the folding and unfolding kinetics of Cys42 have been studied at pH 7, but because the cysteine labeling experiments had to be conducted at a higher pH (pH 8), the global folding and unfolding kinetics have been characterized at pH 8 in the study presented here, using fluorescence as the probe. It should be noted that it was also important to show that the folding kinetics of the four single-cysteine-containing variants are similar, just as their stabilities are similar (Figure 1B).

Like at pH 7, the unfolding reaction of Cys42 at pH 8 can be described as a single-exponential process (Figure 4A) whose amplitude accounts for the entire change in fluorescence during unfolding (Figure 4C). The logarithm of the observed rate

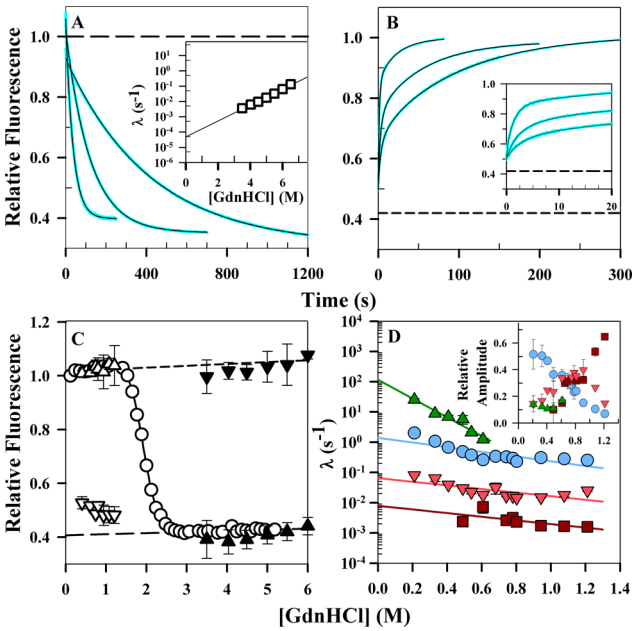


Figure 4. Fluorescence-monitored kinetics of refolding and unfolding of Cys42 at pH 8.0 and 25 °C. (A) Representative kinetic traces of unfolding (solid cyan lines) in 5.5, 4.5, and 3.5 M GdnHCl (left to right, respectively). The solid black lines are fits to a single-exponential equation. The dashed black line represents the signal of the native protein in the absence of GdnHCl. Each of the traces is normalized to a value of 1 for the native protein. The inset shows the rate constants of unfolding at different GdnHCl concentrations. The solid black line is a fit to eq 2. (B) Representative refolding kinetic traces (solid cyan lines) in 0.4, 0.6, and 0.8 M GdnHCl (from left to right, respectively). The solid black lines through the data represent fits to a sum of three exponentials. The dashed line shows the signal of the protein unfolded in 3.5 M GdnHCl, from which refolding commenced. The inset shows the initial part of the kinetic traces. Each of the traces is normalized to a native signal of 1. (C) Kinetic vs equilibrium amplitudes. (○) Equilibrium unfolding transition. The solid line through the data represents a fit to a two-state unfolding model. The extrapolated native and unfolded protein baseline signals are represented by the short dash and long dash lines, respectively. The equilibrium amplitude is compared with the kinetic t = 0 (▽ and ▼) and t = ∞ (△ and ▲) signals for refolding (▽ and △) and unfolding (▼ and ▲). (D) Observed rates of the very fast (green triangles), fast (light blue circles), slow (light red triangles), and very slow (dark red squares) phases of refolding. The inset shows the relative amplitudes of the four phases. The error bars, wherever shown, represent the spread in the measurements from at least two different experiments. The solid lines through the data have been drawn by inspection only.

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

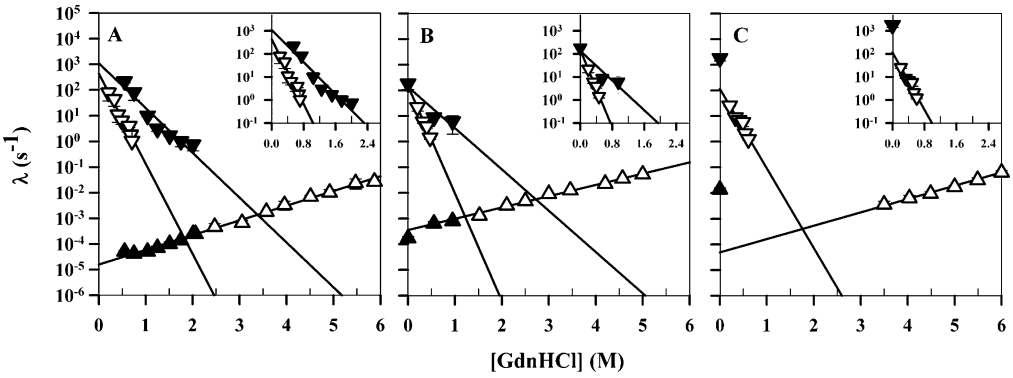


Figure 5. Comparison of the rates monitored by fluorescence and thiol labeling. The observed rates of global unfolding (△) and of the very fast phase of refolding (▽), measured by fluorescence, and the k_{op} (▲) and k_{cl} (▼) rates of each cysteine residue, measured by thiol labeling, are shown for (A) Cys74, (B) Cys63, and (C) Cys42. The insets in each panel show an expanded view of the closing rates and the fluorescence-monitored refolding rates at low denaturant concentrations. The solid lines are fits of the data to eqs 2 (unfolding data) and 3 (refolding data). The error bars represent the spread in the measurements from at least two different experiments.

constant of unfolding increases linearly with an increase in GdnHCl concentration (Figure 4A, inset). While the unfolding rate constants of Cys74 and Cys13 are comparable to those of Cys42, those of Cys63 are 8-fold higher (Figure S3 of the Supporting Information). The increase in the unfolding rate of Cys63 is consistent with a decrease in the native-state stability caused by the mutation. The observed unfolding rates of the four variants have similar dependencies on GdnHCl concentration: a fit of the data to eq 2 indicates that m_{λ_u} has an average value of 1.2 ± 0.05 M^{-1}. Hence, the Tanford β value $[(RTm_{\lambda_u})/m_G]$ is 0.2, indicating that the transition state (TS) for global unfolding at a high GdnHCl concentration has only 20% of the solvent accessibility of the U state.

Like at pH 7, the folding reaction of Cys42 at pH 8 can be described in terms of four observable kinetic phases: very fast, fast, slow, and very slow (Figure 4B). The logarithm of the observed rate constant of each of the four kinetic phases shows a linear denaturant dependence (Figure 4D). The relative amplitude of the very fast phase remains constant at ∼10%, while that of the slow phase increases at the expense of the fast phase (Figure 4D, inset). At GdnHCl concentrations above 0.6 M, the amplitude of the very slow phase increases at the expense of the slow phase. The refolding kinetics of Cys74 and Cys63 are similar to that of Cys42 (Figure S4 of the Supporting Information); the refolding kinetics of Cys13 was not measured because of a lack of kinetic data from SX measurements with which to compare it. Overall, the fluorescence-monitored folding and unfolding measurements, taken in this study at pH 8, are consistent with the refolding mechanism delineated previously at pH 7[35] (see the text of the Supporting Information), not just for Cys42 but also for Cys74 and Cys63 (Figures S3 and S4 of the Supporting Information).

The kinetic and thermodynamic parameters listed in Table 1 show that the mutations do not introduce significant changes into the basic mechanisms of unfolding and refolding.

Comparison of Site-Specific Opening and Closing of Structure to Global Unfolding and Folding. Figure 5 and Table 1 compare the values of k_{op} and k_{cl} at different GdnHCl concentrations to the values of the unfolding rate constant and the very fast apparent folding rate constant, obtained from fluorescence measurements. For both Cys74 and Cys63, the values of k_{op} at low denaturant concentrations agree well with the global unfolding rates, extrapolated to zero denaturant. In the case of Cys74, the values of k_{cl} are at least 2-fold higher than the very fast folding rate, indicating that the burial of C74 takes place faster than any global refolding event. For Cys74 and Cys63, the denaturant dependence of k_{cl} is significantly lower than that of the very fast rate constant. In the case of Cys42, k_{op} and k_{cl} could be compared to the global unfolding and refolding rates in 0 M GdnHCl only (see above); k_{op} is 3000-fold faster than the global unfolding rate at zero denaturant, while k_{cl} is 15-fold faster than the very fast folding rate.

Figure 6 and Table 1 compare the values of ΔG_{op} for each of the cysteine residues to the global free energy of unfolding,

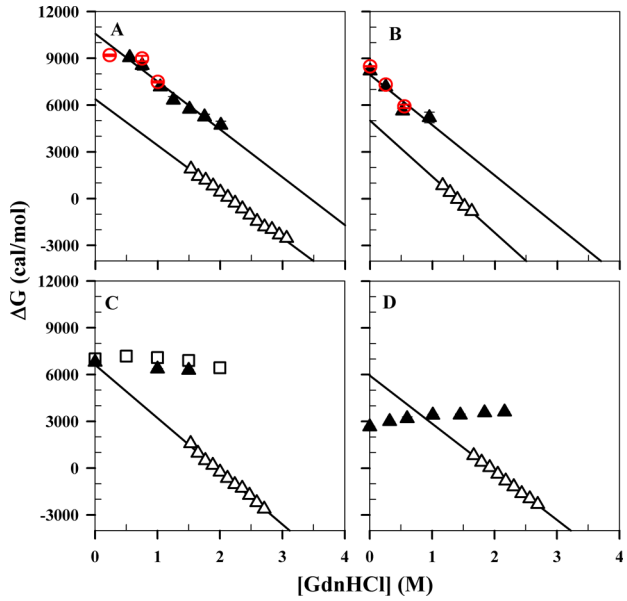


Figure 6. Comparison of the free energies of global unfolding, ΔG_u (△), and the free energies of opening, ΔG_{op} (▲), of the cysteine thiols determined by PDS labeling for (A) Cys74, (B) Cys63, (C) Cys42, and (D) Cys13. For Cys74 and Cys63 MNEI, the ΔG_{op} values were determined from the linear regression fits of the observed rates shown in panels B and D of Figure 3 (red circles) as well as by fitting only the last 30% of the curves shown in panels A and C of Figure 3 to eq 4 (▲). For Cys42, ΔG_{op} values were also measured by DTNB labeling (□). The solid lines through the data are fits to eq 1. The error bars represent the spread in the measurements from at least two different experiments.

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

ΔG_u, determined from fluorescence-monitored equilibrium unfolding transitions. For Cys74 and Cys63, ΔG_{op} was determined from data at low denaturant concentrations (Figure 3B,D), where labeling took place in the SX2 limit, by fitting the data to eq 6 (see above), as well as by fitting only the last 30% of the dependencies of k_{obs} on PDS concentration to eq 4. At the same GdnHCl concentration, identical values were obtained for ΔG_{op} (Figure 6A,B). This suggested that fitting the data at higher denaturant concentrations (>1.25 M for Cys74 and >0.5 M for Cys63), where only the last 30% of the dependence of k_{obs} on PDS concentration could be measured, to eq 4 can also yield reliable values of ΔG_{op}.

For both Cys74 and Cys63, ΔG_{op} is 4.1 and 3.9 kcal mol^{-1}, respectively, higher than ΔG_u at all GdnHCl concentrations studied. For Cys42, in the absence of denaturant, the value of ΔG_{op} is similar to that of ΔG_u, but its value (6.9 kcal mol^{-1}) surprisingly does not change with an increase in GdnHCl concentration up to 2 M. In the case of Cys42, values of ΔG_{op} were obtained from both PDS and DTNB labeling, and Figure 6C shows the good agreement between the two sets of values. In the case of Cys13, ΔG_{op} has a value of 2.7 kcal mol^{-1} in the absence of GdnHCl, and it also shows a negligible dependence on GdnHCl concentration up to 2 M. Measurement of the labeling kinetics at still higher denaturant concentrations (Figure S5 of the Supporting Information) indicates that exposure of the C13 side chain results in very little change in the exposed surface area of the protein, even at 6 M GdnHCl.

The kinetic and thermodynamic parameters determined from thiol labeling and fluorescence measurements for Cys74, Cys63, Cys42, and Cys13 are compared in Table 1.

DISCUSSION

The unfolding of monellin is known to commence via the formation of a DMG, accompanied by a loss of packing interactions between the α-helix and the β-sheet, in the submillisecond time domain.[3] In this study, native-state SX measurements were used to study site-specific opening and closing events of cysteine side chains, at the interface of the α-helix and the β-sheet, to identify structural events associated with the entry of water into interior of the protein, during and immediately following the formation of the DMG. The kinetics of side chain exposure could be compared across the single-cysteine variants because the proteins do not differ significantly in their mechanisms of folding and unfolding (Figure 1 and Figures S3 and S4 of the Supporting Information). The native-state SX measurements of the four single-cysteine-containing variants of monellin have identified three high-energy intermediates in which the cysteine thiols become accessible to labeling by PDS (Figure 7). These intermediates are more unstable than the N state by ~2.7, 6.9, and 10.6 kcal mol^{-1}. The denaturant dependence of ΔG_{op} is a measure of the surface area of the protein that becomes solvent-exposed by the opening event.[18,23] k_{op} and k_{cl} provide information about the intervening energetic barriers and allow the placement of intermediate structures along the unfolding reaction coordinate, with respect to the major unfolding barrier.

Detection of a High-Energy Intermediate after the Major Unfolding Barrier. In the case of Cys74, the agreement between the k_{op} and the extrapolated unfolding rates (Figure 5A) suggests that the thiol becomes exposed in a labeling competent species that forms after the rate-limiting barrier to unfolding at high denaturant concentrations. The exchange competent form appears to be have an energy

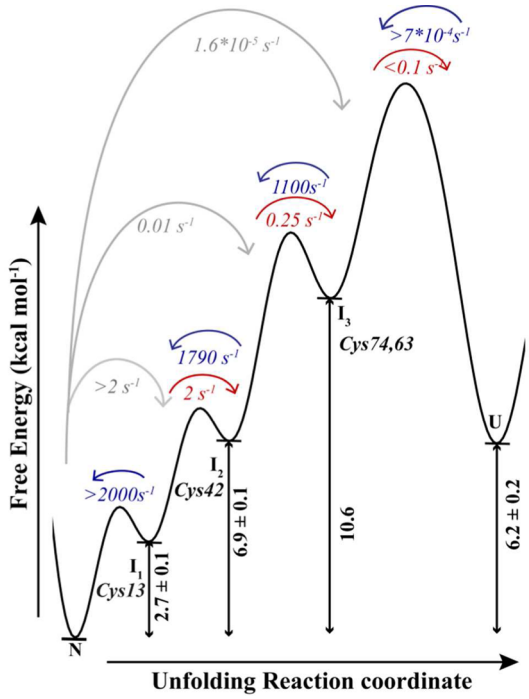


Figure 7. Free energy profile showing the positions of the intermediates detected by thiol labeling along the unfolding reaction coordinate, in 0 M GdnHCl. The energy levels of the unfolded (U) and intermediate (I) states with respect to the native state (N) are colored black. The rates at which the N state undergoes a transition to each of the detected intermediates, which is the opening rate (k_{op}), measured experimentally for each of the variant proteins, are colored gray. The measured closing rates (k_{cl}), colored blue, are the rates at which the thiols in each of the intermediates close back to the N state. The individual forward rates of crossing each of the intervening barriers, calculated from the measured rates (text of the Supporting Information), are colored red.

significantly higher than that of the globally unfolded state (Figure 6A). It should be noted that the values determined for ΔG_{op} at low GdnHCl concentrations (<1 M) are very reliable (see Results). Moreover, very similar values (Figure 6A) were obtained when only the data in the SX2 limit were fit to eq 6, and when only the last 30% of the observed dependence was fit to eq 4 (see Results). Thus, it appears that the values obtained for ΔG_{op} at higher (>1.25 M) GdnHCl concentrations, where only the last 30% of the dependence could be measured and fit to eq 4, are also likely to be reliable.

A higher value of ΔG_{op} compared to ΔG_u indicates "super protection" in the labeling competent species, which can result from an overestimation of the value of ΔG_{op} or have a structural basis. Partial structure in the labeling competent form may lead to the value of k_b, determined for a fully exposed cysteine (see Materials and Methods), being an overestimate, and consequently ΔG_{op} being overestimated. ΔG_{op} should therefore be considered the upper limit for the difference in energy between the labeling competent form and the N state.[44] However, in this case, for the labeling competent species to be similar in energy to the U state, the value of k_b would have to be 200-fold lower than that estimated from the unfolded protein, which is unlikely. The difference between the values of ΔG_{op} and ΔG_u therefore appears to be real. A correction due to proline isomerization, which is only ~1.2 kcal mol^{-1}, in the case of monellin,[35] would also not account for the energy difference

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

of 4.1 kcal mol^{-1} between the U state and the labeling competent species. Such super protection has been interpreted as the presence of partial structure in a high-energy intermediate or residual structure in the U state in previous HX studies.[18,20,46,47] A good agreement between the values of k_b, determined for an unstructured peptide and the unfolded protein, as well as the similar values of k_b determined across all four proteins (Figure S1 of the Supporting Information), eliminates the possibility of a residually structured unfolded state being populated under nativelike conditions. It appears, therefore, that the labeling takes place in a partially structured high-energy intermediate, I_3, which is not part of the unfolded ensemble. The similar denaturant dependences of ΔG_{op} and ΔG_u (Table 1) for Cys74 confirm that I_3 is largely unfolded.

The agreement between k_{op} and the extrapolated unfolding rates (Figure 5A) suggests that I_3 forms after the rate-limiting barrier to unfolding at high denaturant concentrations; but if I_3 were to form after the overall rate-limiting step in unfolding at low denaturant concentrations, the labeling of C74 would occur in U and not in I_3. The observation that the value of ΔG_{op} for the labeling of C74 is significantly higher than the value of ΔG_u suggests that while I_3 forms after the rate-limiting step of unfolding at high denaturant concentrations, it forms before the rate-limiting step of unfolding at low (<1 M) GdnHCl concentrations (Figures 3A,B and 6A). Panels A–C of Figure 8 show the free energy profiles that explain how labeling of C74 takes place at different GdnHCl concentrations. At low GdnHCl concentrations (<0.5 M), where the labeling of C74 in I_3 occurs in the SX2 limit, the barrier separating I_3 from U would necessarily have to be higher than that separating I_3 from N, so that the protein shuttles between N and I_3 and is labeled in I_3.

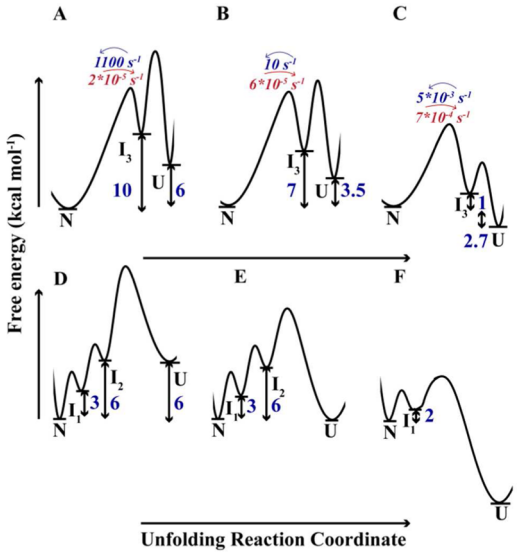


Figure 8. Free energy profiles describing the formation of the intermediate states identified by thiol labeling of Cys74 (A–C) and Cys42 and Cys13 (D–F). The energy levels of the intermediate (I_3) and the unfolded (U) states are shown with respect to the native state (N) in (A) 0 M GdnHCl, in which labeling of the C74 side chain is observed to occur only in the SX2 limit, (B) 1 M GdnHCl, in which C74 labeling takes place in the SX1 and SX2 limits, and (C) 3 M GdnHCl, in which C74 labeling occurs only in the SX1 limit. The relative barriers and energy levels of I_1 and I_2 are shown with respect to N in (D) 0 M GdnHCl, in which the protein is fully folded, (E) 2 M GdnHCl, which is the midpoint of denaturant-induced unfolding of the protein, and (F) 7 M GdnHCl, in which the protein is fully unfolded.

In 1 M GdnHCl (Figure 8B), the labeling of C74 in I_3 occurs in both the SX1 and SX2 limits (Figure 3A). Hence, the rate-limiting barrier to overall unfolding must remain between I_3 and U at this GdnHCl concentration. The value of k_{cl} is lower than at lower GdnHCl concentrations, and labeling in I_3 occurs in the SX2 limit at low PDS concentrations when $k_{cl} > k_b[L]$ and in the SX1 limit at high PDS concentrations when $k_b[L] > k_{cl}$. Kinetic simulations based on the mechanism shown in Scheme 2 (Figure 9A), with I_3 on-pathway to U, were conducted using KINSIM[43] and show that as long as $k_b[L]$ exceeds the rate from I_3 to U, labeling in the SX1 limit takes place predominantly in I_3.

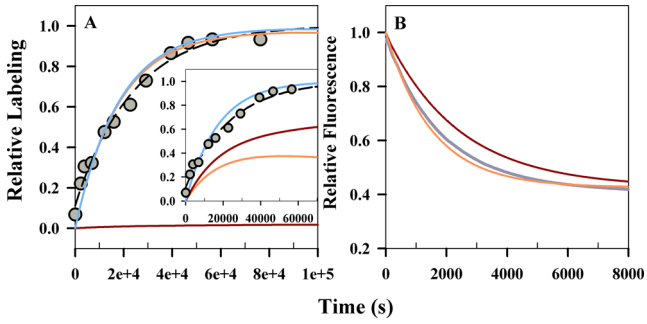


Figure 9. Kinetic simulations for comparison with the fluorescence and SX data for Cys74. (A) The experimental trace of thiol labeling (gray circles) with 3 mM PDS, in 1 M GdnHCl, was compared to the traces simulated according to Scheme 2. The values of k_{op} and k_{cl} in 1 M GdnHCl are 6×10^{-5} and 10 s^{-1}, respectively. The rate of labeling of an exposed cysteine thiol in 1 M GdnHCl (k_3 and k_4) is 80 s^{-1}. The black dashed line is a single-exponential fit through the experimental data. The formation of the I_3·L (orange) and U·L (dark red) species, formed by labeling of I_3 and U, respectively, is shown along with the formation of the total labeled species (light blue) for k_2 set to a value of 1 s^{-1} (main figure) and 100 s^{-1} (inset). When the labeling rate is greater than k_2, labeling takes place primarily from I_3. (B) The experimental trace of fluorescence (gray), in 3 M GdnHCl, was compared to the traces simulated according to Scheme 2, with k_2 set to values of 0.01 s^{-1} (dark red) and 0.1 s^{-1} (orange). The values of k_{op} and k_{cl} in 3 M GdnHCl are 7×10^{-4} and 0.01 s^{-1}, respectively. The fluorescence signal of I_3 was fixed at a value similar to that of U. The simulated traces with k_2 values of ≥0.1 s^{-1} compare well with the experimental data.

At still higher GdnHCl concentrations (Figure 8C), the barrier separating I_3 from U appears to become lower than that separating I_3 from N. The TS separating N and I_3 has only 20% more exposed surface area than N, while that separating I_3 and U is essentially U-like in its exposed surface area because I_3 is U-like (see above). Hence, increasing the GdnHCl concentration stabilizes the latter TS more than it stabilizes the former TS. Figure 8 shows that I_3 lies on the native side of the rate-limiting barrier at low GdnHCl concentrations but on the unfolded side of the rate-limiting barrier at high denaturant concentrations, because the rate-limiting barrier changes with an increase in GdnHCl concentration. Kinetic simulations of the fluorescence data based on the mechanism shown in Scheme 2 (Figure 9B) show that the simulated traces fit the experimental data well only when the I_3 to U rate exceeds the backward rate from I_3 to N. The simulations also suggest that most of the fluorescence change occurs between N and I_3, which is consistent with the observation that I_3 is largely

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608–3620

unfolded. It should be noted that at high GdnHCl concentrations, the rate-limiting step of unfolding is the barrier between N and I_3, and that the values obtained for k_{op} at low GdnHCl concentrations correspond to unfolding over only this barrier; hence, the values obtained for λ_U at high GdnHCl concentrations using fluorescence measurements extrapolate linearly to the values obtained for k_{op} at low GdnHCl concentrations using SX measurements (Figure 5A). In 3 M GdnHCl, k_{cl} is low enough that even at the lowest PDS concentration of 25 μM, k_b[L] exceeds k_{cl}; consequently, the labeling in I_3 occurs only in the SX1 limit.

ΔG_{op} values could be reliably determined in the SX2 limit (Figure 3D), below 0.5 M GdnHCl, for Cys63, as well. The difference in the values of ΔG_{op} and ΔG_u for Cys63 (3.9 kcal mol^{-1}) is similar to that for Cys74 (4.1 kcal mol^{-1}), indicating that both the thiols become labeled in the same intermediate, I_3. This is also consistent with the observation that for both Cys63 and Cys74, k_{op} and λ_u agree with each other (Figure 5B). The destabilization of the native state, by 1.8 kcal mol^{-1} (Figure 1B), results in the value of ΔG_{op} of Cys63 (8.3 kcal mol^{-1}) being lower than that of Cys74 (10.6 kcal mol^{-1}). However, an 8-fold increase in the unfolding rate of Cys63 (Table 1), instead of the expected 20-fold increase, suggests that the mutation destabilizes the transition state for unfolding, as well. This explains the lower k_{cl} for Cys63 (183 s^{-1}) compared to that of Cys74 (1100 s^{-1}) (Table 1). The lower denaturant dependence of k_{cl} compared to that of the very fast refolding phase further confirms that I_3 is on-pathway between the N and U states.

Detection of Intermediates on the Native Side of the Barrier. In the case of Cys42, k_{op} is faster than global unfolding (Figure 5C), indicating that a labeling competent intermediate, I_2, is formed much faster than the U state and before the major rate-limiting step of unfolding. The intermediate can be placed at an energy level similar to that of the unfolded state, not only in 0 M GdnHCl (Figure 6C) but also in ≤2 M GdnHCl. This observation indicates that very little change in surface area occurs upon unfolding of N to I_2. This is not surprising because I_2 precedes the TS for unfolding, and the TS itself has only 20% of the additional surface area that U has, compared to that of N (see Results).

In the case of Cys13, the labeling competent species is formed faster than 2 s^{-1} (Figure 3F and Table 1), indicating that the labeling takes place in a species, I_1, formed before the major unfolding barrier, at an energy level of 2.7 kcal mol^{-1}, compared to the N state (Figure 6D). I_1 has surface exposure similar to that of the N state, and hence, the free energy change upon formation of I_1 from N (ΔG_{op}) has little dependence on GdnHCl concentration (Figure 6D and Figure S5 of the Supporting Information). Because the major unfolding reaction is much slower ($\lambda_u^{H_2O} \sim 10^{-5}$ s^{-1}) than the formation of I_1, it appears that this intermediate also precedes the TS of unfolding. The observation that at >1 M GdnHCl the value of ΔG_{op} is higher than that of ΔG_u suggests that I_1, like I_2, remains populated even at high denaturant concentrations.

Because of the high barrier to unfolding in the case of monellin, the protein remains trapped in intermediates I_1 and I_2 long enough for the labeling to take place from the intermediates, and not from the U state (Figure 8D–F). Although the possibility that I_1 and I_2 are off-pathway to the U state cannot be ruled out because of the lack of availability of kinetic data, the lack of any dependence on GdnHCl concentration of the free energies of unfolding of N to I_1 and I_2, together with the observation that the TS itself is very nativelike in its solvent exposure, suggests that they might be on-pathway to U. A similar structural transition observed in the on-pathway DMG intermediate detected in previous studies suggests that I_1 is also an on-pathway productive species.[3] It should be noted, however, that monellin is known to unfold via at least two routes,[28,35,37] and it is possible that I_1 and I_2 are populated on competing parallel pathways.

Estimation of Intermediate Populations and Barrier Heights. In previous studies, HX NMR[18,21,44] and relaxation dispersion NMR methods[17,48−52] have been successful in providing high-resolution structural information about transiently populated intermediates. However, HX measurements monitor only the amide backbone and are inherently limited by the dynamic range of the ΔG and m values.[19,53] On the other hand, NMR methods fail to detect species that are more than 3 kcal mol^{-1} higher in energy than the ground state. In this study of native-state SX of monellin, high-energy intermediates have been detected over a wide range of stabilities between 3 and 10 kcal mol^{-1}, with respect to the native state (Table 1). This has been possible because of rapid rates of labeling and the slow unfolding rate of monellin due to a high unfolding free energy barrier. The populations of each of the intermediates, estimated from the values of ΔG_{op}, are 10^{-6}% for Cys74 and Cys63, 8 × 10^{-4}% for Cys42, and 1% for Cys13 (see the Supporting Information). SX therefore serves as a very powerful technique for the detection of intermediates populated to extremely low levels, even when the N state constitutes >99% of the total population under nativelike conditions.

An intermediate can be detected by SX as long as its transition to another state is slow compared to the labeling reaction. Given the low populations of the detected intermediates, it becomes important to confirm that these high-energy states, in each of which one or more cysteine thiols become solvent-exposed, are populated long enough for the labeling to occur. The experimentally measured values of k_{op}, k_{cl}, k_b, and ΔG_{op} for Cys74, Cys63, Cys42, and Cys13, which are listed in Table 1, have been used to quantify the barriers (Figure 7) among the N, I, and U states, as discussed in the text of the Supporting Information. These calculations suggest that, irrespective of the connectivity between the detected intermediates, the measured labeling kinetics are consistent with the labeling of the cysteine thiols in the high-energy intermediates shown in Figure 7.

Structural Transitions Identified by Thiol Labeling. The solvent-accessible surface area (SASA) of each of the residues, chosen for this study, is ∼1 Å^2 in the native structure. As shown in Figure 10, the removal of contacting residues that are within 4−5 Å results in the SASA becoming comparable to that in the extended conformation of an Ala-X-Ala tripeptide, which is used as a model of the unfolded state.[54] The fact that the solvent exposure of a cysteine thiol requires residues in the vicinity to also lose side chain packing interactions and move apart therefore allows thiol labeling to monitor structural changes in a nonlocal manner even though it probes only a single residue at a time. In this study, site-specific thiol labeling has identified opening events of varying amplitudes, in different parts of the protein structure. While the opening of the C13 and C42 side chains does not expose any significant surface area across a large range of denaturant concentrations (Figure 6C,D), that of the C74 and C63 residues is accompanied by a surface area change that is comparable to the global unfolding of the protein (Figure 6A,B).

dx.doi.org/10.1021/bi401493t | Biochemistry 2014, 53, 3608−3620

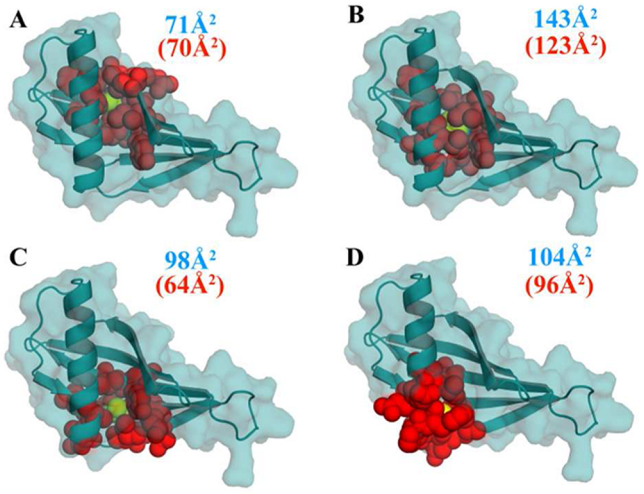


Figure 10. Structural arrangements around the cysteine side chain, which provide protection against exposure to the surrounding solvent. Each cysteine residue is shown as a yellow sphere, and the neighboring residues within 5 Å are shown as red spheres for (A) Cys74, (B) Cys63, (C) Cys42, and (D) Cys13. The structures have been drawn from the PDB entry 1IV7 using PyMOL. The SASAs of each side chain probed, after removal of neighboring residues within 5 Å, are reported (red). These are compared to the SASAs of the residues in an Ala-X-Ala tripeptide (blue).

The side chains probed in this study monitor the helix−sheet interface in monellin. It is likely that the labeling of Cys13 reports on the movement of the single α-helix, because the packing of the C13 residue against the β-sheet appears to be responsible for its burial (Figure 10D). It is possible that the labeling of Cys74, Cys63, and Cys42 might be monitoring the dissolution of the individual β-strands. If this interpretation is correct, then the 10^5-fold dispersion in the k_{op} values (Table 1) allows the corresponding structural events to be temporally resolved in a hypothetical structural model for unfolding. The labeling of the C13 side chain is the fastest [>2 s^{-1} (Figure 3F)], suggesting a rapid movement of the helix away from the β-sheet, accompanied by a negligible change in surface area (Figure 6D). Even though C42 is present on the adjacent β-strand and packs against C13, it gets exposed at a 100-fold slower rate (Figure 5C). The slower labeling rates of Cys74 and Cys63 further confirm that the packing interactions in β-strands are maintained for much longer and dissolve only after the major barrier to unfolding has been crossed (Figure 5A,B).

Correlation to Structural Transitions in the DMG. The detection of intermediates on the native side of the major barrier allows direct correlations to be made to the DMG, which is known to precede the rate-limiting step of unfolding.[1] The formation of the DMG in monellin is accompanied by the fraying of the α-helix away from the β-sheet.[3] This study also suggests that the movement of the helix away from the β-strands is one of the earliest structural transitions leading to increased solvent accessibility of C13. It would appear that intermediate I_1, in which C13 in the helix becomes competent to exchange, is either the DMG itself or formed immediately after the formation of the DMG. If I_1 is the DMG, it would mean that the DMG is solvated, at least in the part of the helix in which C13 resides. Further studies are needed to confirm such an interpretation.

Disruption of side chain packing interactions, which are known to stabilize the structure of the native state, accompanies the formation of the DMG. However, most of the studies that have detected a DMG have probed residues on, or close to, the surface of the protein.[55] Although the side chains probed in this work have comparable SASA values in the native state, they are present at different depths within the protein, thereby providing a description of the extent of packing interactions throughout the molecule. The minimal depth of a residue in monellin is 3.1 Å. C63 and C74 side chains, having depths of 8.3 and 9.5 Å, respectively, are the deepest residues and therefore appear to constitute the core, which is found to remain protected until most of the protein has unfolded. C42 and C13, which are at depths of 7.2 and 5.4 Å, respectively, are closer to the surface and are exposed faster. Thus, these data suggest that side chain packing in the core of the protein might be maintained in the DMG and is disrupted only after the rate-limiting unfolding barrier, in intermediate I_3. I_3 appears to resemble a wet molten globule (WMG) in which the core of the protein is hydrated.

The formation of the dry globule is also accompanied by a change in the distance between W4 and Q29,[3] which are both present close to the surface of the protein. This suggests that while core packing might be retained, surface interactions are disrupted in the DMG. A recent multiprobe study of the protein barstar has also shown that the core of the protein remains "dry" in the DMG and becomes solvated only in the wet globule.[5]

CONCLUSIONS

The observation of a DMG, which has a dry interior, on the unfolding pathway of proteins has kindled considerable interest in how different parts of the structure become accessible to the surrounding solvent.[10,55,56] This study has provided an understanding of the nature and time scales of the structural events that lead to the exposure of buried side chains in the protein. Site-specific SX measurements have been used to directly monitor the loss of packing interactions in the protein, subsequent to the formation of the DMG. It has been shown that side chain packing interactions are lost incrementally, in multiple steps, resulting in high-energy, low-population intermediates on both sides of the major unfolding barrier. This suggests that solvent exposure takes place noncooperatively and that some parts of the protein continue to remain protected until most of the interactions maintaining the native structure are lost. Further studies are required to correlate the changes in secondary structure, which is retained in the DMG and the WMG, with those in the tertiary structure, which is already disrupted. This will allow the detected intermediates to be identified more accurately with the dry and wet globules. Finally, because the unfolding studies have been conducted under nativelike conditions, they shed direct light on the mechanism of protein folding that will be protein unfolding in reverse under identical conditions.

ASSOCIATED CONTENT

Supporting Information

A detailed description of fluorescence measurements and calculations of SASA and depths and analysis of free energies and rates to determine barrier heights. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Fax: 91-80-23636662. Phone: 080-2366-6150. E-mail: jayant@ncbs.res.in.

Funding

This work was funded by the Tata Institute of Fundamental Research and by the Department of Science and Technology, Government of India. J.B.U. is the recipient of a J. C. Bose National Research Fellowship from the Government of India.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Shachi Gosavi, Hemanth Giri Rao, and members of our laboratory for discussions.

■ ABBREVIATIONS

MNEI, single-chain monellin; GdnHCl, guanidine hydrochloride; DTT, dithiothreitol; PDS, 4,4′-dipyridyl disulfide; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); SX, thiol labeling; HX, hydrogen exchange; C_m, midpoint of the denaturant-induced unfolding transition; PDB, Protein Data Bank.

■ REFERENCES

(1) Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature 375*, 513−515.

(2) Hoeltzli, S. D., and Frieden, C. (1995) Stopped-flow NMR spectroscopy: Real-time unfolding studies of 6-^{19}F-tryptophan-labeled *Escherichia coli* dihydrofolate reductase. *Proc. Natl. Acad. Sci. U.S.A. 92*, 9318−9322.

(3) 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.

(4) Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc. Natl. Acad. Sci. U.S.A. 107*, 4955−4960.

(5) Sarkar, S. S., Udgaonkar, J. B., and Krishnamoorthy, G. (2013) Unfolding of a small protein proceeds through dry and wet globules and a solvated transition state. *Biophys. J. 105*, 2392−2402.

(6) Jha, S. K., and Marqusee, S. (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc. Natl. Acad. Sci. U.S.A. 111*, 4856−4861.

(7) Dasgupta, A., Udgaonkar, J. B., and Das, P. (2014) Multistage Unfolding of a SH3 Domain: An Initial Urea-Filled Dry Molten Globule Precedes a Wet Molten Globule with Non-Native Structure. *J. Phys. Chem. B*, DOI: 10.1021/jp410019f.

(8) Finkelstein, A. V., and Shakhnovich, E. I. (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers 28*, 1681−1694.

(9) Shakhnovich, E. I., and Finkelstein, A. V. (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers 28*, 1667−1680.

(10) Baldwin, R. L., Frieden, C., and Rose, G. D. (2010) Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins 78*, 2725−2737.

(11) 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.

(12) Zitzewitz, J. A., and Matthews, C. R. (1993) Protein engineering strategies in examining protein folding intermediates. *Curr. Opin. Struct. Biol. 3*, 594−600.

(13) Nath, U., and Udgaonkar, J. B. (1995) Perturbation of a tertiary hydrogen bond in barstar by mutagenesis of the sole His residue to Gln leads to accumulation of at least one equilibrium folding intermediate. *Biochemistry 34*, 1702−1713.

(14) Feng, H., Takei, J., Lipsitz, R., Tjandra, N., and Bai, Y. (2003) Specific non-native hydrophobic interactions in a hidden folding intermediate: Implications for protein folding. *Biochemistry 42*, 12461−12465.

(15) Pradeep, L., and Udgaonkar, J. B. (2002) Differential Salt-induced Stabilization of Structure in the Initial Folding Intermediate Ensemble of Barstar. *J. Mol. Biol. 324*, 331−347.

(16) Dasgupta, A., and Udgaonkar, J. B. (2012) Four-state folding of a SH3 domain: Salt-induced modulation of the stabilities of the intermediates and native state. *Biochemistry 51*, 4723−4734.

(17) Korzhnev, D. M., and Kay, L. E. (2008) Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Acc. Chem. Res. 41*, 442−451.

(18) Bai, Y., Sosnick, T., Mayne, L., and Englander, S. (1995) Protein folding intermediates: Native-state hydrogen exchange. *Science 269*, 192−197.

(19) Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Mol. Biol. 3*, 782−787.

(20) Bhuyan, A. K., and Udgaonkar, J. B. (1998) Two structural subdomains of barstar detected by rapid mixing NMR measurement of amide hydrogen exchange. *Proteins 30*, 295−308.

(21) Englander, S. W. (2000) Protein folding intermediates and pathways studied by protein folding. *Annu. Rev. Biophys. Biomol. Struct. 29*, 213−238.

(22) Juneja, J., and 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.

(23) 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.

(24) Silverman, J. A., and Harbury, P. B. (2002) Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J. Biol. Chem. 277*, 30968−30975.

(25) 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.

(26) 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.

(27) Stratton, M. M., Cutler, T. A., Ha, J. H., and Loh, S. N. (2010) Probing local structural fluctuations in myoglobin by size-dependent thiol-disulfide exchange. *Protein Sci. 19*, 1587−1594.

(28) Jha, S. K., Dasgupta, A., Malhotra, P., and Udgaonkar, J. B. (2011) Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry. *Biochemistry 50*, 3062−3074.

(29) 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.

(30) Silverman, J. A., and Harbury, P. B. (2002) The equilibrium unfolding pathway of a $(\beta/\alpha)_8$ barrel. *J. Mol. Biol. 324*, 1031−1040.

(31) Tancredi, T., Iijima, H., Saviano, G., Amodeo, P., and 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.

(32) Morris, J. A., and Cagan, R. H. (1972) Purification of monellin, the sweet principle of *Dioscoreophyllum cumminsii*. *Biochim. Biophys. Acta 261*, 114−122.

(33) Ogata, C., Hatada, M., Tomlinson, G., Shin, W.-C., and Kim, S.-H. (1987) Crystal structure of the intensely sweet protein monellin. *Nature 328*, 739−742.

(34) Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T., Akiyama, S., and Fujisawa, T. (2005) Specific collapse followed by slow hydrogen-bond formation of β-sheet in the folding of single-chain monellin. *Proc. Natl. Acad. Sci. U.S.A. 102*, 2748−2753.

(35) 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.
(36) Kimura, T., Maeda, A., Nishiguchi, S., Ishimori, K., Morishima, I., Konno, T., Goto, Y., and Takahashi, S. (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. U.S.A. 105*, 13391−13396.
(37) Aghera, N., and Udgaonkar, J. B. (2013) The Utilization of Competing Unfolding Pathways of Monellin Is Dictated by Enthalpic Barriers. *Biochemistry 52*, 5770−5779.
(38) Ellman, G. L. (1959) Tissue sulfhydryl groups. *Arch. Biochem. Biophys. 82*, 70−77.
(39) Grassetti, D. R., and Murray, J. F., Jr. (1967) Determination of sulfhydryl groups with 2,2′- or 4,4′-dithiodipyridine. *Arch. Biochem. Biophys. 119*, 41−49.
(40) Riener, C. K., Kada, G., and Gruber, H. J. (2002) Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4′-dithiodipyridine. *Anal. Bioanal. Chem. 373*, 266−276.
(41) Agashe, V. R., and 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.
(42) Hvidt, A., and Nielsen, S. O. (1966) Hydrogen Exchange in Proteins. In *Advances in Protein Chemistry* (Anfinsen, C. B., Edsall, J. T., and Richards, F. M., Eds.) Advances in Protein Chemistry, Vol. *21*, pp 287−386, Elsevier, Amsterdam.
(43) Barshop, B. A., Wrenn, R. F., and Frieden, C. (1983) Analysis of numerical methods for computer simulation of kinetic processes: Development of KINSIM—A flexible, portable system. *Anal. Biochem. 130*, 134−145.
(44) Wildes, D., and Marqusee, S. (2004) Hydrogen-exchange strategies applied to energetics of intermediate processes in protein folding. *Methods Enzymol. 380*, 328−349.
(45) Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins 17*, 75−86.
(46) Grantcharova, V. P., and Baker, D. (1997) Folding dynamics of the src SH3 domain. *Biochemistry 36*, 15685−15692.
(47) Nicholson, E. M., Mo, H., Prusiner, S. B., Cohen, F. E., and Marqusee, S. (2002) Differences between the prion protein and its homolog Doppel: A partially structured state with implications for scrapie formation. *J. Mol. Biol. 316*, 807−815.
(48) Di Nardo, A. A., Davidson, A. R., Dobson, C. M., Kay, L. E., Korzhnev, D. M., Salvatella, X., and Vendruscolo, M. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature 430*, 586−590.
(49) Tollinger, M., Kloiber, K., Agoston, B., Dorigoni, C., Lichtenecker, R., Schmid, W., and Konrat, R. (2006) An isolated helix persists in a sparsely populated form of KIX under native conditions. *Biochemistry 45*, 8885−8893.
(50) Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R., and Kay, L. E. (2010) A transient and low-populated protein-folding intermediate at atomic resolution. *Science 329*, 1312−1306.
(51) Bouvignies, G., Vallurupalli, P., Hansen, D. F., Correia, B. E., Lange, O., Bah, A., Vernon, R. M., Dahlquist, F. W., Baker, D., and Kay, L. E. (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature 477*, 111−114.
(52) Hansen, A. L., Bouvignies, G., and Kay, L. E. (2013) Probing slowly exchanging protein systems via $^{13}C\alpha$-CEST: Monitoring folding of the Im7 protein. *J. Biomol. NMR 55*, 279−289.
(53) Bai, Y., and Englander, S. W. (1996) Future directions in folding: The multi-state nature of protein structure. *Proteins: Struct., Funct., Genet. 24*, 145−151.
(54) Richards, F. M. (1977) Areas, Volumes, Packing, and Protein Structure. *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
(55) Bhattacharyya, S., and Varadarajan, R. (2013) Packing in molten globules and native states. *Curr. Opin. Struct. Biol. 23*, 11−21.
(56) Bernstein, R., Schmidt, K. L., Harbury, P. B., and Marqusee, S. (2011) Structural and kinetic mapping of side-chain exposure onto the protein energy landscape. *Proc. Natl. Acad. Sci. U.S.A. 108*, 10532−10537.