THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 52, pp. 37479–37491, December 28, 2007

Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry*

Received for publication, August 13, 2007, and in revised form, October 22, 2007 Published, JBC Papers in Press, October 24, 2007, DOI 10.1074/jbc.M706714200

Santosh Kumar Jha[1] **and Jayant B. Udgaonkar**[2]

From the National Centre for Biological Sciences, Tata Institute of Fundamental Research, Gandhi Krishi Vignana Kendra (GKVK) Campus, Bangalore 560065, India

It has been difficult to obtain directly residue-specific information on side chain packing during a fast (ms) protein folding reaction. Such information is necessary to determine the extent to which structural changes in different parts of the protein molecule are coupled together in defining the cooperativity of the overall folding transition. In this study, structural changes occurring during the major fast folding reaction of the small protein barstar have been characterized at the level of individual residue side chains. A pulsed cysteine labeling methodology has been employed in conjunction with mass spectrometry. This provides, with ms temporal resolution, direct information on structure formation at 10 different locations in barstar during its folding. Cysteine residues located on the surface of native barstar, at four different positions, remain fully solvent-accessible throughout the folding process, indicating the absence of any ephemeral nonnative structure in which these four cysteine residues get transiently buried. For buried cysteine residues, the rates of the change in cysteine-thiol accessibility to rapid chemical labeling by the thiol reagent methyl methanethiosulfonate appear to be dependent upon the location of the cysteine residue in the protein and are different from the rate measured by the change in tryptophan fluorescence. But the rates vary over only a 3-fold range. Nevertheless, a comparison of the kinetics of the change in accessibility of the cysteine 3 thiol with those of the change in the fluorescence of tryptophan 53, as well as of their denaturant dependences, indicates that the major folding reaction comprises more than one step.

To obtain an understanding of the cooperativity of the structural transitions accompanying the folding of unfolded protein to its unique native fold has been the central objective of many protein folding studies (1–4). Many protein folding reactions have been described as cooperative two-state U ⇌ N transitions, implying that native structure forms in a concerted all-or-none manner, with the formation of one contact facilitating the formation of many others (3, 5). There is, however, a growing body of work suggesting that protein folding/unfolding transitions may be highly noncooperative and even be gradual structural transitions (6–12). For example, time-resolved anisotropy and fluorescence resonance energy transfer experiments, as well as NMR experiments, have shown that structure is lost incrementally during the denaturant-induced equilibrium unfolding of barstar (6, 9, 12). High resolution probes, such as UV resonance Raman spectroscopy and NMR, indicate that the equilibrium unfolding of Trp-cage (10), GCN4-like leucine zipper (7), CHABII (8), and BBL (11) is spatially decoupled and occurs in many steps. Nearly all evidence for gradual or multistep folding comes from equilibrium unfolding studies, where high resolution structural probes can be used easily. Kinetic evidence is needed (13, 14), especially evidence that provides structural information at the individual residue level, but this has been limited so far (15–20). The major question that remains unaddressed is whether different regions of a protein form a structure in a synchronized or in an unsynchronized manner, during the major folding reaction of any protein.

Pulsed cysteine labeling (SX)[3] provides direct structural information on the fate of individual residues during the folding process and has been shown to be an excellent probe for studying structure formation during the fast folding/unfolding reactions of several proteins at the level of individual side chains (21–25). In brief, side chains located in different parts of the protein structure are mutated to cysteine, one at a time, and the solvent accessibility of the individual cysteine thiol group to rapid chemical labeling is measured at different times of folding. The extent to which a particular cysteine residue is involved in structure formation at any time of refolding is reflected by the fraction of molecules in which the cysteine thiol gets labeled at that time. In this study, the pulsed SX methodology has been coupled with mass spectrometry for the first time to explore the cooperativity of the refolding reaction of barstar.

The folding pathway of the small protein barstar has been characterized extensively (26–31), and under strongly stabilizing conditions it can be represented as follows.

$$U \underset{1\ ms}{\rightleftharpoons} I_E \xrightarrow{100\ ms} I_L \xrightarrow{100\ s} N$$

SCHEME 1

* This work was supported by the Tata Institute of Fundamental Research and by the Department of Biotechnology, Government of India. All mass spectra were collected at the Mass Spectrometry Facility at the National Centre for Biological Sciences. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Recipient of a Shyama Prasad Mukherjee Fellowship awarded by the Council of Scientific and Industrial Research, India.

[2] To whom correspondence should be addressed. Fax: 91-80-23636862; E-mail: jayant@ncbs.res.in.

[3] The abbreviations used are: SX, cysteine labeling; MMTS, methyl methanethiosulfonate; ESI, electrospray ionization.

where I_E represents a highly heterogeneous early intermediate, consisting of different structural forms, and is shown to be populated during the initial few milliseconds of refolding (28, 30, 31). A rapid equilibrium between U and I_E is established before the major structural transition to the late intermediate I_L occurs. I_L has also been shown to be a heterogeneous ensemble of intermediates (32–34). Different members of the I_E ensemble are populated in different solvent conditions; hence, the structural properties of I_E appear different in different folding conditions (30, 31). The extent of cooperativity present in the major structural transition (the I_E to I_L structural transition), which constitutes the fast phase of folding, is poorly understood at the individual residue level. Barstar offers itself as an interesting model system for this type of study, because the kinetics of the fast phase (*i.e.* of the I_E to I_L transition) are different when monitored using different probes, under some but not all folding conditions (29, 31).

In this study, the pulsed SX methodology in conjunction with mass spectrometry has been applied to a library of 10 single Cys, single Trp-containing mutant forms of barstar, in which the single Cys residue is located at various buried and exposed locations of the protein. Fig. 1*a* shows the locations of the single tryptophan and the cysteine residues so introduced. For each mutant protein, the accessibility of the individual cysteine thiol group to a cysteine-specific labeling reagent, methyl methanethiosulfonate (MMTS), and the fluorescence of the sole tryptophan were measured at different times during the folding process. Only a 3-fold dispersion is seen in the site-specific rates of the fast change in cysteine thiol accessibility during folding, but a more detailed analysis of the dependence on urea concentration of the kinetics of the fast folding reaction measured at two different sites (Cys^3 and Trp^{53}) in one of the mutant proteins suggests that the fast folding reaction comprises more than one step.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The method for the purification of barstar and its mutant variants has been described in detail previously (36). Ten different mutant variants, Cys3, Cys14, Cys25, Cys36, Cys40, Cys42, Cys62, Cys67, Cys82, and Cys89, each with a single Trp residue (Trp^{53}) and a single Cys residue (at the residue position indicated in the name), were generated by site-directed mutagenesis (18, 19). Protein purity was checked by mass spectrometry using a Micromass Q-TOF Ultima mass spectrometer coupled with an ESI source. The masses determined for Cys3, Cys14, Cys25, Cys36, Cys40, Cys42, Cys62, Cys67, Cys82, and Cys89 were 10,232, 10,216, 10,232, 10,232, 10,232, 10,202, 10,190, 10,232, 10,246, and 10,216 Da, respectively. These masses indicate that the N-terminal methionine residue remained uncleaved during the expression of the mutant proteins. Protein concentrations were determined for all of the above mutant proteins by measuring the absorbance at 280 nm, using an extinction coefficient of 10,000 M^{-1} cm^{-1}.

Reagents and Chemicals

Boric acid, EDTA disodium salt, MMTS, 5,5′-dithiobis(2-nitribenzoic acid), and cysteine·HCl was of ultrapure grade from Sigma. Urea (ultrapure grade) was from U. S. Biochemical Corp. Dithiothreitol (ultrapure grade) was obtained from Invitrogen, formic acid (GPR grade) was from BDH, and acetonitrile (HPLC grade) was from Qualigens.

Buffers and Solutions

The native buffer used for all of the equilibrium and kinetic experiments was composed of 200 mM sodium borate and 1 mM EDTA at pH 9.2. The unfolding buffer was native buffer containing 6 M urea for all the kinetic experiments or 8 M urea for all of the equilibrium unfolding experiments at pH 9.2. Urea concentrations were determined from the measurement of the refractive index on an Abbe 3L refractometer from Milton Roy. All buffers and solutions were filtered through 0.22-μm filters and degassed before use. All of the experiments were carried out at 25 °C.

Preparation of MMTS-labeled Protein

MMTS-labeled protein was prepared by reaction of the protein in 8 M urea (unfolding buffer) at pH 9.2, with a 100-fold molar excess of MMTS for ~5 min. The labeling reaction was quenched by the addition of a 1000-fold molar excess (to the protein) of cysteine·HCl (in 1% formic acid) to the reaction mixture. The addition of cysteine·HCl also decreased the pH of the solution to ~2, which ensured that labeled protein did not lose any label. Following this, the labeled protein was separated from cysteine, urea, and other small molecules present in the reaction mixture by passing the protein through a Hi-Trap Sephadex G-25 desalting column on an Akta chromatography system. The extent of labeling was checked by mass spectrometry, and the protein was found to be >95% labeled as judged from the expected 46-Da increase in the mass of the protein.

Measurement of the Stability of MMTS

The stability of MMTS, the labeling reagent used in this study, depends upon the pH of the solution in which it is reconstituted; hence, its decomposition process imposes a time constraint on the use of a MMTS solution after its preparation. The rate of decomposition of MMTS was measured under various conditions (*i.e.* at pH 9.5 and 8, in water at pH 6 ± 0.5, and in aqueous solutions containing 0.6–2.0 M urea at pH 6.3 ± 0.5, as described earlier) (37). In brief, fresh solutions of dithiothreitol and 5,5′-dithiobis(2-nitribenzoic acid) were prepared in 20 mM Tris buffer at pH 8. Dithiothreitol was added to the solution containing excess 5,5′-dithiobis(2-nitribenzoic acid), and the concentration of the released TNB^{2-} ions was determined by measuring the absorbance at 412 nm, using an extinction coefficient of 14,100 M^{-1} cm^{-1}. The concentration of MMTS in solution was determined as a function of time by monitoring the loss of absorbance of the TNB^{2-} ions at 412 nm after the addition of an aliquot of the solution to the reaction mixture containing dithiothreitol and excess 5,5′-dithiobis(2-nitribenzoic acid). MMTS was found to decompose with a time constant of ~4 min at pH 9.5 and ~60 min at pH 8 (data not shown). It is, however, fairly stable when reconstituted in water and aqueous solutions containing urea. It was found to decompose less than 5% in 3 h when reconstituted in aqueous solution containing 2.0 M urea at pH 6.5 ± 0.3 (data not shown). For all

of the pulsed SX experiments, the MMTS solution was reconstituted in water containing 0.6–2.0 M urea at pH 6.3 ± 0.5 and was used within 2 h of its preparation.

Kinetics of Change in Cysteine Accessibility during Refolding

All pulsed SX experiments were carried out using a Biologic SFM-400 Q/S unit operating in the quenched flow mode. Both refolding and labeling reactions were performed at pH 9.2 and 25 °C. The protein was unfolded in 6 M urea (unfolding buffer) for at least 3 h prior to refolding experiments. After a variable time 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.

Three different quenched flow programs were used to achieve the refolding times of 0, 5–23, and >23 ms. For the 0 ms refolding time point, 30 μl of 140.4 mM MMTS (in water) were mixed with 330 μl of refolding buffer inside the quenched flow machine for 5 ms, and the resulting solution was mixed with 40 μl of unfolded protein solution (150 μM stock) for 4 ms. The labeling reaction was quenched with the addition of 103 μl of 410 mM cysteine·HCl (in 1% formic acid) solution. For refolding times in the range of 5–23 ms, refolding was initiated by mixing 24 μl of unfolded protein solution (195 μM stock) with 208 μl of refolding buffer at pH 9.2 and 25 °C in various delay loops to achieve the desired refolding time points, and then the resulting solution was pulsed with 80 μl of 40.9 mM MMTS solution (in water containing 0.6 M urea) for 4 ms. The labeling reaction was quenched with the addition of 80 μl of 410 mM cysteine·HCl (in 1% formic acid) solution. To achieve refolding times greater than 23 ms, refolding was initiated by mixing 12 μl of unfolded protein solution (195 μM stock) with 104 μl of refolding buffer, at pH 9.2 and 25 °C in a 90-μl delay loop (total intermixer volume = 116 μl). After a variable delay time, a pulse consisting of 40 μl of 40.9 mM MMTS (in water containing 0.6 M urea) was applied for 4 ms. The labeling reaction was quenched with the addition of 40 μl of 410 mM cysteine·HCl (in 1% formic acid) solution.

The concentration of protein at the time of labeling was 15 μM in all of the above pulsed SX experiments. The concentration of MMTS at the time of labeling was 10.5 mM, except in the experiments where the dependence on MMTS concentration of the cysteine accessibility-monitored refolding kinetics was studied. In those experiments, an identical protocol was followed except that the calculated amount of MMTS solution (9.75 M stock solution) was dissolved in water containing 0.6 M urea, so as to give the desired MMTS concentration at the time of labeling. The concentration of cysteine·HCl in all of the above experiments was 10-fold higher than that of MMTS at the time of quenching.

In the experiments where the dependence on urea concentration of the cysteine accessibility monitored refolding kinetics was studied, a protocol identical to that described above was followed, except that the refolding buffer contained the calculated amount of urea, so as to give the desired urea concentration at the time of refolding. Also, the MMTS solution was reconstituted in water containing the desired urea concentration.

Control experiments were done to ensure that the concentrations of MMTS and cysteine·HCl used in the above experiments 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 MMTS solution.

Processing of Samples for Mass Spectrometry

All samples were processed in an identical manner. Each sample was desalted on an Akta chromatography system, using a Hi-Trap Sephadex G-25 desalting column. MilliQ water at pH 3 (pH adjusted with formic acid) was used for elution. A control experiment was performed to ensure that there was no cross-contamination between two samples during desalting. Five samples collected at the same time point of refolding, when desalted one after the other, gave the same composition of labeled and unlabeled protein (within ±3%), as determined by mass spectrometry (data not shown).

A high concentration of free cysteine was present in the samples after the pulsed SX experiments. It was conceivable that the labeled protein might be reduced if the samples were not desalted for a long time. The following control experiment was done to determine the time frame within which the samples had to be desalted. Samples were collected in duplicate after the pulsed SX experiments. One set (Set A) was desalted right after the pulsed SX experiment (within 2 h), and another set (Set B) was desalted 15 h later. The amount of labeled protein in each sample of Set B is found to be only 1–2% less than the corresponding sample of Set A. Moreover, no peaks corresponding to the addition of the cysteine moiety to the unlabeled protein were observed in any of the mass spectra in any of the above experiments. To be safe, desalting was completed within 2 h after the pulsed SX experiments for all samples.

Determination of the Extent of Labeling by ESI-Mass Spectrometry

The extents of labeling in samples from the pulsed SX experiments were determined using ESI-mass spectrometry. A Micromass Q-TOF Ultima mass spectrometer coupled with an ESI source, which was operated under Mass Lynx software control, was used. For acquisition of the mass spectra, 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. 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 μl/min. All of the spectra were collected in the positive ion mode. The concentration of protein in each sample was typically 2–3 μM, and typically an ion count of ~150 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 two protein species (unlabeled and MMTS-labeled protein) was observed in each 1-s scan. For each sample, the data acquired over 60 s were averaged. All of the resultant m/z spectra were processed in the following way using the Mass Lynx version 4.0 software. Background noise subtraction was done using a second order poly-

nomial below 30% of the curve with a tolerance value of 0.01, followed by a two-point smoothening with a Savitzky-Golay algorithm (supplied with the Mass Lynx software) using a smoothening window (in channels) of ±23. The extent of labeling was determined from these smoothened m/z spectra by calculating the average relative ion intensity of the labeled protein from the ninth, tenth, and eleventh charged state peaks (these were the three most intense peaks in the mass spectra).

Fluorescence-monitored Equilibrium and Kinetic Folding Experiments

Folding was monitored using the change in fluorescence of Trp^{53} as a probe. All equilibrium unfolding experiments were performed on a Fluoromax-3 fluorimeter (Jobin Yvon). The protein samples were incubated in different concentrations of urea for at least 3 h prior to the fluorescence measurements. Identical results were obtained if the time of incubation was 24 h. Excitation of tryptophan fluorescence was carried out at 295 nm, using a slit width of 0.5 nm. Emission was monitored at 320 nm using a slit width of 10 nm. The final protein concentrations in equilibrium unfolding experiments were 3–5 μM.

All kinetic experiments were carried out using a Biologic SFM-4 stopped-flow machine. Proteins were unfolded in 6 M urea (unfolding buffer) for at least 3 h prior to the refolding experiments. Refolding was initiated by mixing 30 μl of unfolded protein with 270 μl of refolding buffer inside the stopped-flow mixing module. Sample excitation was carried out at 295 nm, and emission was monitored at 320 nm using an Oriel bandpass filter with a bandwidth of ±10 nm. In all 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 concentrations in fluorescence-monitored kinetic refolding experiments were 15–25 μM. In the experiments where the dependence of the fluorescence-monitored refolding kinetics on urea concentration was studied, an identical protocol as above was followed, except that refolding was initiated by appropriate dilution of refolding buffer, unfolding buffer, and unfolded protein inside the stopped-flow mixing module, so as to give the final desired urea concentration at the time of refolding.

Data Analysis

Determination of the Bimolecular Rate Constants for MMTS Labeling of a Cysteine Thiol in a Protein—The exchange reaction (SX) between a thiol labeling reagent and a protected thiol group of a protein can be modeled by a Linderstrom-Lang type of equation (22, 24), which was conceptualized originally to explain amide-hydrogen exchange phenomena in proteins. The exchange reaction between a thiol labeling reagent and a protected thiol group of a protein can be modeled as follows.

$$\text{Closed(-S-H)} \underset{k_{closed}}{\overset{k_{open}}{\rightleftharpoons}} \text{Open(-S-H)} \xrightarrow[[\text{MMTS}]]{k_b} \text{Exchanged(-S-S-CH}_3\text{)}$$

SCHEME 2

A cysteine-thiol protected in the protein structure can get labeled with MMTS only when a structural opening reaction (*i.e.* local or global unfolding) exposes that thiol transiently to the solvent. In the above scheme, k_{open} and k_{closed} are the kinetic rate constants for opening and closing of a cysteine thiol residue in the closed-to-open reaction, and k_b is the second order rate constant of the reaction of that thiol group with MMTS in the unfolded protein. Under steady state conditions, the observed rate constant of exchange of the thiol in the closed state is given by the following.

$$k_{ex} = \frac{k_{open} \times k_b[\text{MMTS}]}{k_{closed} + k_b[\text{MMTS}]} \quad \text{(Eq. 1)}$$

Two limiting cases of Equation 1 exist, depending upon the relative rates of the closing reaction (k_{closed}) and of chemical exchange from the open unfolded state (k_b[MMTS]).

If $k_{closed} \lll k_b[\text{MMTS}]$, then the following is true.

$$k_{ex} = k_{open} \quad \text{(Eq. 2)}$$

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

On the other hand, if $k_{closed} \ggg k_b[\text{MMTS}]$, then the following is true.

$$k_{ex} = \frac{k_{open} \times k_b[\text{MMTS}]}{k_{closed}} = K_{open} \times k_b[\text{MMTS}] \quad \text{(Eq. 3)}$$

This is known as the SX2 limit, and under this condition, k_{ex} measures the equilibrium constant between the closed and open states. The above two mechanisms can be distinguished, because the rate of labeling for the SX2 mechanism is dependent on the concentration of MMTS, whereas it is not for the SX1 mechanism.

Analysis of the Equilibrium Unfolding Data—All equilibrium unfolding transition curves were analyzed using a two state N ⇌ U model (38). Raw data were converted into fraction unfolded *versus* [urea] plots, and the two-state fit to the data yielded the values for the free energy change for unfolding, ΔG_U, and the midpoint of the unfolding transition (C_m).

Analysis of the 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 three exponentials,

$$A(t) = A_\infty - A_F e^{-\lambda_F t} - A_I e^{-\lambda_I t} - A_S e^{-\lambda_S t} \quad \text{(Eq. 4)}$$

where $A(t)$ and A_∞ are the observed amplitudes at times t and infinity, respectively; λ_F, λ_I, and λ_S are the apparent rate constants of the fast, intermediate, and slow phases; and A_F, A_I, and A_S are the respective amplitudes.

The relative amplitude of the fast phase was determined using the following equation,

$$\alpha_F = \frac{A_F}{(S_U - S_N)} \times 100 \quad \text{(Eq. 5)}$$

where α_F represents the relative amplitude of the fast phase, A_F is the observed amplitude of the fast phase in 0.6 M urea, and S_U and S_N are the fluorescence signals of the unfolded protein in 6 M urea and of the native protein in 0.6 M urea, respectively.

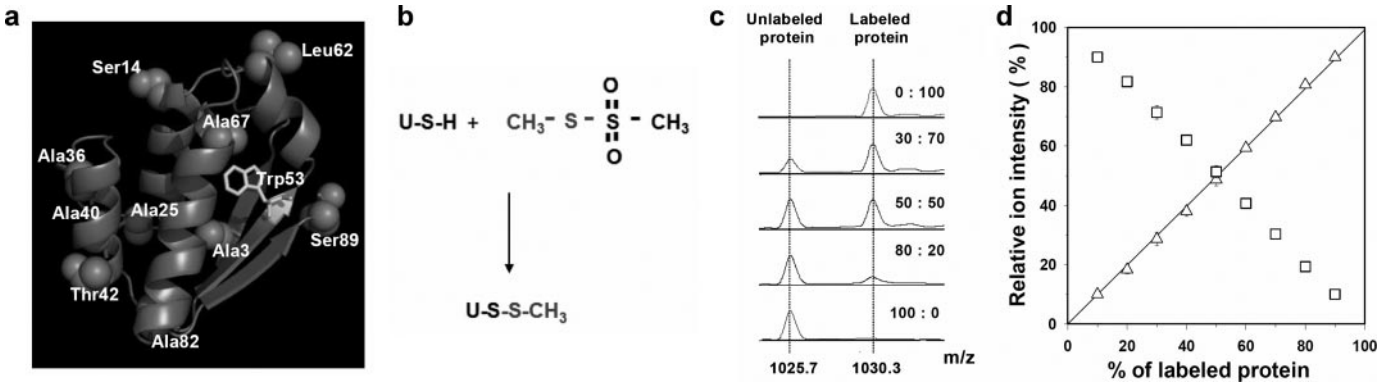

FIGURE 1. **Quantitative estimation of the extent of the labeling of a cysteine thiol in barstar.** *a*, the location of Trp[53] and the residues that were replaced by cysteine residues in the single Cys, single Trp-containing mutant proteins are shown in the structure of barstar (Protein Data Bank accession code 1A19), which has been drawn using the program PyMOL (35). *b*, MMTS reacts with a solvent-exposed thiol group of a protein, increasing the mass of the protein by 46 Da. *c*, mass spectra (10th charged state) of mixtures of unlabeled and MMTS-labeled Cys82 mutant protein, which had been mixed in the molar ratios indicated. *d*, the relative populations of unlabeled (□) and fully labeled (△) Cys82 molecules in a mixture of the two, determined from the their relative ion intensities calculated using the ninth, tenth, and eleventh charged states in the mass spectra, scale linearly with the molar ratio with which the labeled and unlabeled molecules were mixed. *d*, the *error bars* represent the S.D. values from three separate experiments.

Analysis of Cysteine Burial-monitored Kinetic Data—The analysis of the pulsed SX data makes use of earlier observations that I_E is populated within a few ms of the commencement of refolding (18, 26–31, 34), and that the transition from I_L to N is very slow (26–28, 32–34). Hence, only the $I_E \rightarrow I_L$ reaction needs to be considered when folding in the 2–300-ms time window is studied. Since the cysteine-labeling pulses were applied only at times in the range of 2–300 ms within the commencement of refolding, only the pulse labeling of I_E and I_L needs to be considered as shown.

$$\mathrm{U} \xrightarrow{<1\ \mathrm{ms}} \mathrm{I_E} \xrightarrow{k_f} \mathrm{I_L} \xrightarrow{100\ \mathrm{s}} \mathrm{N}$$

$$\downarrow k_E \qquad \downarrow k_L$$

$$\mathrm{I_E\text{-}L} \qquad \mathrm{I_L\text{-}L}$$

SCHEME 3

k_f is the rate of burial of any particular cysteine thiol during the folding transition from I_E to I_L. Protein labeled either as I_E-L or I_L-L will be measured as the MMTS-labeled protein in the mass spectrum. k_E and k_L are the apparent, pseudo-first order rate constants for the labeling of a particular cysteine thiol in I_E and I_L, respectively, under the labeling conditions used. The values of k_E and k_L will depend on the protection against labeling offered by the structures of I_E and I_L, respectively. In the strongly stabilizing conditions used to carry out the folding reaction, all three folding transitions, $U \rightarrow I_E$, $I_E \rightarrow I_L$, and $I_L \rightarrow N$, can be assumed to be irreversible, because the reverse unfolding reactions will be slow in comparison (26–28).

If the burial of any particular cysteine thiol occurs in a single step during the $I_E \rightarrow I_L$ reaction, then the fraction of molecules that will get labeled by the 4 ms MMTS pulse at different times of folding, t, is given by the following.

$$A_F(t) = ((e^{-k_f t}) \times (1 - e^{-0.004 k_E}) + (1 - e^{-k_f t}) \times (1 - e^{-0.004 k_L})) \times 100 \quad \text{(Eq. 6)}$$

The first term on the right-hand side of Equation 6 indicates the molecules of I_E that get labeled, and the second term indicates the molecules of I_L that get labeled by the 4 ms labeling pulse.

The relative amplitude of the observed fast phase in the cysteine labeling experiments was determined using the equation,

$$\alpha'_F = \frac{(A_F(0) - A_F(\infty))}{(S'_U - S'_N)} \times 100 \quad \text{(Eq. 7)}$$

where α'_F is the relative amplitude of the fast phase; $A_F(0)$ and $A_F(\infty)$ are the values of $A_F(t)$ (in Equation 6) at time 0 and infinity, and S'_U and S'_N are the extent of labeling of unfolded protein in 6 M urea and of native protein in 0.6 M urea, respectively.

Calculation of Reduced Amplitudes—For fluorescence-monitored kinetics, the reduced amplitude (α) of the fast phase at any urea concentration is calculated using the following equation,

$$\alpha = \left(\frac{(S_U - S_F)}{(S_U - S_N)}\right) \times 100 \quad \text{(Eq. 8)}$$

where S_F represents the fluorescence signal observed at the end of the fast phase at that particular urea concentration, and S_U and S_N are the fluorescence signals of the unfolded protein in 6 M urea and of the native protein in 0.6 M urea, respectively.

In the case of cysteine accessibility-monitored kinetics, the reduced amplitude (α') at any urea concentration is calculated using the following equation,

$$\alpha' = \left(\frac{(S'_U - S'_F)}{(S'_U - S'_N)}\right) \times 100 \quad \text{(Eq. 9)}$$

where S'_F is the extent of labeling observed at the end of the fast phase at that particular urea concentration, and S'_U and S'_N are the extent of labeling of unfolded protein in 6 M urea and of native protein in 0.6 M urea, respectively.

RESULTS

ESI-Mass Spectrometry Estimates Quantitatively the Relative Amounts of Labeled and Unlabeled Proteins in a Mixture of the Two—Fig. 1*b* shows the chemical reaction between a solvent-

exposed thiol group of a protein and the cysteine-specific labeling reagent, MMTS, which is known to label a solvent-exposed cysteine thiol very rapidly (22, 37). MMTS transfers one -S-CH_3 moiety to the thiol group of the protein 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 proteins shows two peaks, 46 Da apart in absolute mass, corresponding to the masses of the two proteins. In order to estimate the relative amount of labeled and unlabeled proteins present in a mixture of the two, the labeled and unlabeled proteins were mixed in different known ratios and fed into the mass spectrometer. Fig. 1*c* shows the mass spectra of the five samples in which the unlabeled and MMTS-labeled Cys82 mutant proteins were mixed in the molar ratio indicated. Fig. 1*d* shows that the relative ion intensities of the labeled and unlabeled proteins predict, within error limits of 5%, the correct molar ratio in which the two proteins were mixed. Hence, ESI-mass spectrometry can be used to estimate quantitatively the relative amounts of labeled and unlabeled proteins in a mixture of the two.

Establishing the Conditions for Pulsed Cysteine Labeling—In order to set the conditions to be used for the labeling pulse, it was necessary to determine how fast MMTS labels a cysteine thiol in a protein when the thiol is fully exposed to the solvent (*i.e.* to know the value of k_b in Equation 3). The pseudo-first order chemical exchange rate constants of the reaction between MMTS and a cysteine thiol that is fully exposed to solvent were determined at various concentrations of MMTS, using urea-unfolded Cys82 (Fig. 2*a*). These apparent first order rate constants were dependent linearly on the concentration of MMTS (Fig. 2*b*), and the slope of the straight line yielded a second order rate constant for the reaction between MMTS and a cysteine thiol that is fully exposed to solvent, as in urea-unfolded Cys82, of 4.8×10^5 M^{-1} s^{-1}, at pH 9.2 and 25 °C. Based on this measurement, the concentration of MMTS in the labeling pulse to be used in the pulsed SX experiments was chosen to be 10.5 mM, so that a 4 ms pulse would label unfolded protein molecules completely. This duration of the labeling pulse is more than 10-fold lower than the time constant of the fast folding reaction.

To determine the extent to which a cysteine thiol is protected in the native protein and to determine the type of SX mechanism, native state exchange was measured as a function of the concentration of MMTS. Fig. 2*c* shows the determination of pseudo-first order chemical exchange rate constants of the reaction between MMTS and the cysteine thiol of native Cys82 at various concentrations of MMTS at pH 9.2 and 25 °C. The plot of log k_{obs} *versus* log[MMTS] was linear over the range 0.15–1.5 mM MMTS with a slope of 1.1, which is within error of the expected value of 1 (data not shown), indicating that SX occurs mainly by the SX2 mechanism. The extent to which the native structure protects the thiol group against labeling can therefore be expressed as the protection factor ($P_{SX} = 1/k_{open}$) (see Equation 3), where P_{SX} is given by the ratio of the second order rate constant of labeling of unfolded protein molecules to the second order rate constant of labeling of native protein molecules. Fig. 2*d* shows the linear dependence of the apparent first order rate constant of labeling of native Cys82 on the concentration of MMTS. The slope of the straight line yields a second order rate constant of 1.7×10^3 M^{-1} s^{-1}, at pH 9.2 and 25 °C. Hence, the cysteine thiol is ~280 times more protected in native Cys82 than in the unfolded protein.

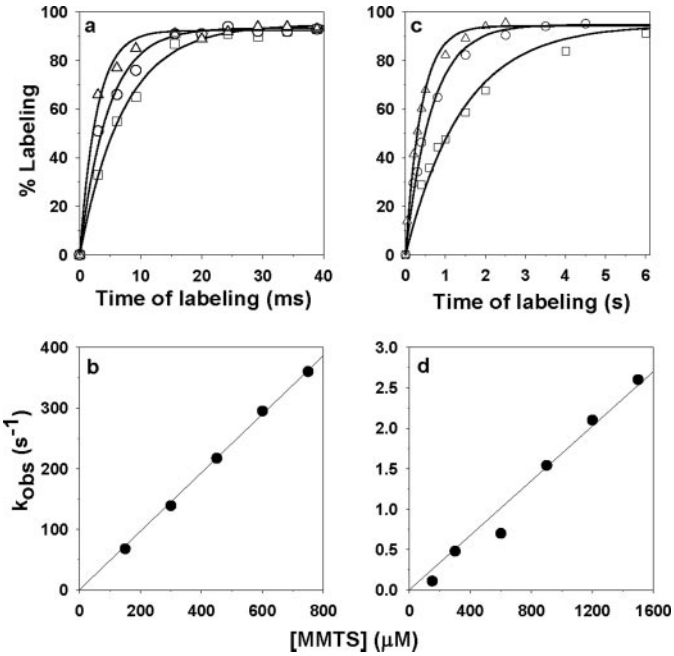

FIGURE 2. **Determination of the extent to which native structure protects against labeling by MMTS at pH 9.2, 25 °C.** *a*, kinetics of labeling of unfolded Cys82 in 8 M urea by MMTS. 15 μM protein in 8 M urea was labeled with 300 μM (□), 450 μM (○), and 750 μM (△) MMTS, and the extent of labeling was determined at different times of reaction. The *solid lines* through the data are least-squares fits to a single exponential equation. *b*, dependence of the rate of labeling of unfolded protein (●) on MMTS concentration. The *solid straight line* represents a linear fit of the data and yields a value for the bimolecular rate constant of 4.8×10^5 M^{-1} s^{-1}. *c*, kinetics of labeling of native Cys82 by MMTS. 15 μM protein in refolding buffer was labeled with 600 (□), 900 (○), and 1500 (△) μM MMTS, and the extent of labeling was determined at different times of reaction. The *solid lines* through the data are least-squares fits to a single exponential equation. *d*, dependence of the rate of labeling of native protein (●) on MMTS concentration. The *solid straight line* represents a linear fit of the data and yields a value for the bimolecular rate constant of 1.7×10^3 M^{-1} s^{-1}.

Fluorescence-monitored Kinetic and Equilibrium Folding Transitions—All 10 mutant proteins have been shown to preserve the structure and function of the wild-type protein (18, 19). Fig. 3 shows the Trp^{53} fluorescence-monitored kinetics of folding in 0.6 M urea, as well as the equilibrium unfolding transitions, of all 10 mutant proteins at pH 9.2. All of the mutant proteins have similar stabilities: the free energy of unfolding is 3.9 ± 0.5 kcal mol^{-1}, across the set of mutant proteins. The midpoints for urea-induced unfolding and the free energies of unfolding are listed in Table 1. In previous studies, these proteins were shown to have similar stabilities at pH 8, whether determined from fluorescence measurements as in the current study (19) or from far-UV circular dichroism measurements (9, 24, 34, 39). The refolding kinetics of the mutant proteins are also similar. The observed rate constants of the fast phase of refolding and their relative amplitudes are listed in Table 1 (see Equations 4 and 5 for the calculation of rate constant and relative amplitude). For each of these mutant proteins, the kinetics of refolding in 0.6 M urea show the following features: 1) there is a 5–15% unobservable burst phase change in fluorescence, which is complete within the dead time of the mixing instrument; 2) the observable kinetics fit to a sum of three-exponen-

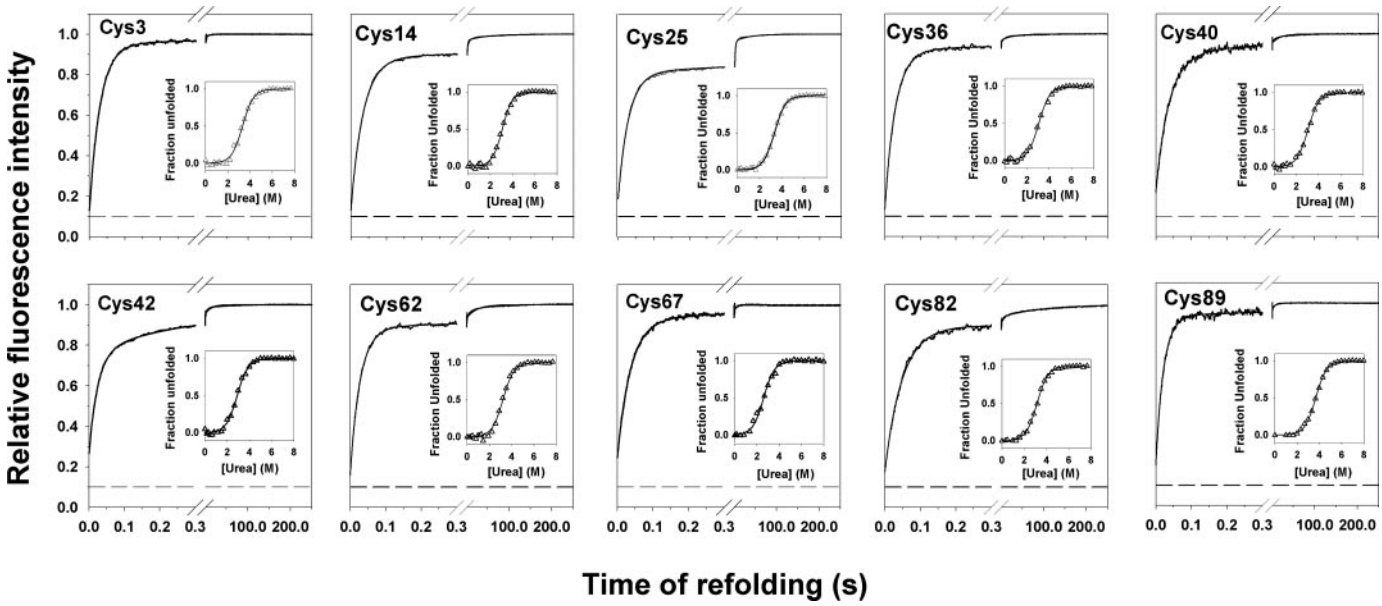

FIGURE 3. **Kinetic and equilibrium folding transitions of the single cysteine-containing variants at pH 9.2.** In each *panel*, the fluorescence-monitored kinetics of refolding in 0.6 M urea are shown for the mutant variant indicated. Each kinetic trace is normalized to a value of 1 for the fluorescence of the native protein in 0.6 M urea, and the *dashed line* represents the relative fluorescence of unfolded protein in 6 M urea. The *solid line* through the kinetic data represents a fit to Equation 4. The *inset* in each *panel* shows the equilibrium folding transition for the mutant protein variant. The fraction folded is plotted against urea concentration.

TABLE 1
Thermodynamics and kinetics of refolding of the different mutant variants of barstar

Single cysteine mutant forms of barstar	Free energy of unfolding (ΔG_U)	Midpoint of unfolding (C_m)	Fluorescence-monitored kinetics		Cysteine burial-monitored kinetics	
			Observed rate[a]	Relative amplitude[b]	Observed rate[c]	Relative amplitude[d]
	kcal mol^{-1}	*M*	s^{-1}	%	s^{-1}	%
Cys3	4.0 ± 0.4	3.3 ± 0.05	34.0 ± 1.0	85 ± 4	52 ± 3.5	78 ± 2
Cys14	3.7 ± 0.5	3.1 ± 0.06	28.5 ± 1.0	83 ± 2	No process	No process
Cys25	4.0 ± 0.5	3.3 ± 0.05	34.0 ± 1.0	85 ± 5	No process	No process
Cys36	3.7 ± 0.3	3.1 ± 0.03	33.5 ± 0.5	85 ± 2	No process	No process
Cys40	3.7 ± 0.4	3.1 ± 0.05	30.0 ± 1.5	80 ± 5	63 ± 3.0	70 ± 3
Cys42	3.6 ± 0.4	3.0 ± 0.05	39.5 ± 1.0	75 ± 5	No process	No process
Cys62	3.9 ± 0.3	3.2 ± 0.04	35.0 ± 1.0	81 ± 1	34 ± 1.0	68 ± 5
Cys67	3.4 ± 0.6	2.8 ± 0.1	34.0 ± 1.0	78 ± 4	52 ± 2.0	84 ± 2
Cys82	3.8 ± 0.5	3.2 ± 0.06	19.5 ± 1.0	83 ± 3	60 ± 4.0	72 ± 2
Cys89	4.4 ± 0.6	3.6 ± 0.1	47.5 ± 1.5	85 ± 3	120 ± 9.0	94 ± 2

[a] Determined using Equation 4.
[b] Determined using Equation 5.
[c] Determined using Equation 6.
[d] Determined using Equation 7.

tials with 75–85% of the total fluorescence change occurring in the fast phase of refolding; and 3) the mutant proteins fold with similar apparent rate constants for the fast phase of refolding ($32 \pm 3.0\ s^{-1}$). It therefore appears that the folding mechanism of barstar is conserved among the set of mutant proteins. Consequently, their folding kinetics can be compared directly with each other. Moreover, the stabilities as well as the folding kinetics of the mutant proteins at pH 9.2 are very similar to those measured earlier at pH 7 or 8 (18, 19, 26–31, 34, 40), indicating that the folding mechanism is the same in the pH range 7–9.2.

Cysteine Accessibility-monitored Refolding Kinetics—Fig. 4 shows the refolding kinetics of the mutant proteins in 0.6 M urea, as monitored by change in solvent accessibility of the single thiol group during refolding, along with the measured solvent accessibilities in the unfolded and native state. In all cases, solvent accessibility was measured by determining the extent of labeling by a 4 ms labeling pulse of 10.5 mM MMTS at pH 9.2.

For the folding of all of the mutant proteins, the extent of labeling levels off at 300 ms of refolding (Fig. 4). This fast phase of folding is also complete at 300 ms when monitored by the change in Trp^{53} fluorescence (Fig. 3), indicating that I_L is completely formed at this time of folding. Fig. 4 shows that the protection factor for labeling of a cysteine thiol in I_L is position-specific, as is also seen for N. In the case of Cys62, Cys67, or Cys89, the protection factor in I_L appears to be the same as that in the native protein; the thiol is labeled to about the same extent in I_L and N. In the case of Cys3, Cys40, and Cys82, the thiol is labeled to a greater extent in I_L than in N. For all of the other proteins (Cys14, Cys25, Cys36, and Cys42), the thiol is labeled to the same extent in I_L, N, and U.

Equation 6 was used to extract a reliable value for k_ρ the rate of burial of a cysteine thiol during the I_E to I_L reaction, from the kinetic data. In using Equation 6 to extract the value of k_ρ any possible exchange into I_E as well as I_L gets taken into account. Importantly, any effect of local chemistry, structural heteroge-

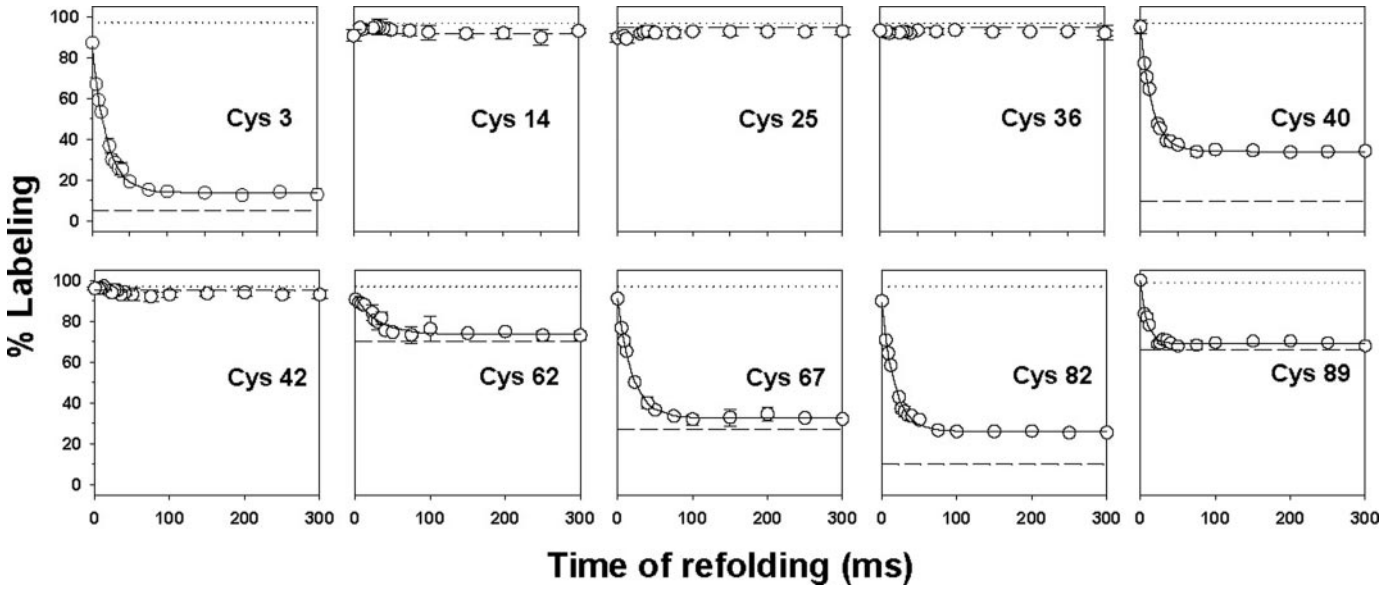

FIGURE 4. **Kinetics of the change in cysteine accessibility during refolding in 0.6 M urea at pH 9.2.** Fractional labeling (○) is plotted against the time of application of the labeling pulse after commencement of refolding, for each of the mutant protein variants, as indicated. In each *panel*, the *solid line* is a fit of the labeling data to Equation 6. The values of k_f, k_E, and k_L extracted from the fits are 52, 484, and 37 s^{-1} for Cys3; 64, 762, and 104 s^{-1} for Cys40; 34, 642, and 332 s^{-1} for Cys62; 52, 600, and 99 s^{-1} for Cys67; 60, 553, and 76 for Cys82; and 119, 1425, and 293 s^{-1} for Cys89, respectively. The *dotted and dashed lines* represent the labeling of unfolded protein in 6 M urea and native protein in 0.6 M urea, respectively. The *error bars* represent the S.D. values of measurements from three different experiments.

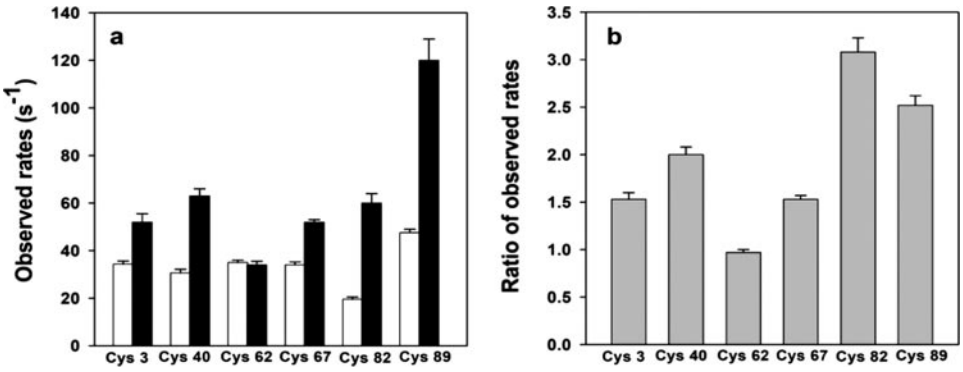

FIGURE 5. **Comparison of the fluorescence and cysteine accessibility-monitored apparent rate constants of fast refolding in 0.6 M urea at pH 9.2.** *a*, the observed fluorescence-monitored (*empty bars*) and cysteine accessibility-monitored (*filled bars*) refolding rate constants for the indicated mutant proteins. *Error bars*, S.D. values from three separate experiments. *b*, ratio of the cysteine accessibility-monitored rate constant to the fluorescence-monitored rate constant. *Error bars*, S.E. values in the determination of ratios.

neity, or local unfolding at a cysteine thiol site, gets reflected in the values of k_E and k_L and not in the value of k_f (see "Discussion"). Values obtained for k_f, as well as for k_E and k_L, the effective rate constants for labeling of the thiol in I_E and I_L during the 4 ms pulse obtained for Cys3, Cys40, Cys62, Cys67, Cys82, and Cys89, each of which contain the single thiol group at fully or partially buried locations, as judged from their measured native state solvent accessibilities, are listed in the legend to Fig. 4. The observed rate constants and the relative amplitudes (see Equation 7 for the calculation of relative amplitudes) of the fast phase of refolding, for each mutant protein, are also listed in Table 1.

There appears to be a 3-fold dispersion in the rates of burial of the six cysteine thiols at different structural locations and in the rate of Trp^{53} fluorescence change, as the protein transits from I_E to I_L (Figs. 4 and 5). The dispersion in rates is small, and although the errors in the determination of the rates are much smaller (see Table 1), the significance of the dispersion must be treated with caution (see "Discussion").

The site-specific values obtained for k_E, the apparent pseudo-first order rate constant of labeling of a particular cysteine thiol in I_E, vary over a 3-fold range (see the legend to Fig. 4). Possible reasons are that local chemistry is different at each site (22) or that folding is carried out in 0.6 M urea, where thiol labeling rates are faster than in 6 M urea (24). Alternatively, the structure of I_E may afford different levels of protection against labeling, within a 3-fold range, to cysteine thiols at different locations on the protein. Measurement of the native state solvent accessibility of the cysteine thiol in Cys14, Cys25, Cys36, and Cys42 indicates that these cysteine thiol groups are completely exposed to solvent in the folded protein (Fig. 4). No transient burial of the cysteine thiol is observed for any of these mutant proteins during the process of refolding; the thiol remains fully solvent-accessible throughout the folding process.

Dependence of Cysteine Accessibility-monitored Refolding Kinetics on MMTS Concentration—When the strength of the labeling pulse is increased by increasing the concentration of MMTS present during labeling from 4–10.5 to 40 mM, for two of the mutant proteins, Cys3 and Cys67, no effect is seen on the rates of cysteine burial (Fig. 6, *a* and *b*).

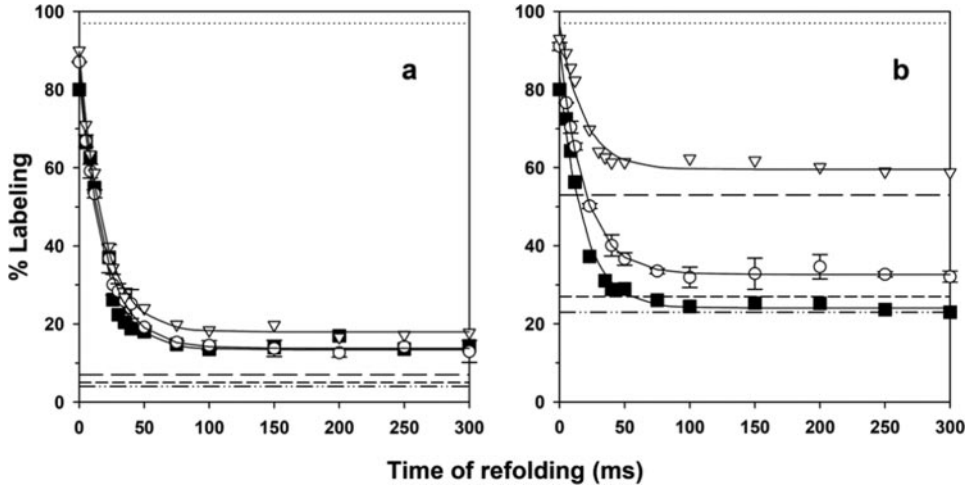

FIGURE 6. **Dependence of the cysteine accessibility-monitored kinetics on MMTS concentration.** Folding was carried out in 0.6 M urea at pH 9.2 and 25 °C. *a*, Cys3; *b*, Cys67. In each *panel*, fractional labeling determined for three different concentrations of MMTS (4 mM (■), 10.5 mM (○), and 40 mM (▽) MMTS) in the labeling pulse is plotted against the time of application of the labeling pulse after commencement of refolding. The *solid line* is a fit of the labeling/refolding data to Equation 6. The relative amplitude of the observed fast phase was determined using Equation 7. The observed rates (k_f) and relative amplitudes of the fast phase of refolding for Cys3 are 55 s^{-1} and 76% for 4 mM MMTS, 52 ± 3.5 s^{-1} and 78 ± 2% for 10.5 mM MMTS, and 54 s^{-1} and 80% for 40 mM MMTS. The observed rates (k_f) and relative amplitudes of the fast phase of refolding for Cys67 are 54 s^{-1} and 85% for 4 mM MMTS, 52 ± 2 s^{-1} and 84 ± 2% for 10.5 mM MMTS, and 54 s^{-1} and 84% for 40 mM MMTS. The *dotted line* represents the labeling of unfolded protein in 6 M urea upon incubation with 4, 10.5, and 40 mM MMTS, respectively, for 4 ms. The *dashed and dotted*, the *short dashed*, and the *long dashed lines* represent the extent of labeling of native protein in 0.6 M urea, upon incubation with 4, 10.5, and 40 mM MMTS, respectively, for 4 ms.

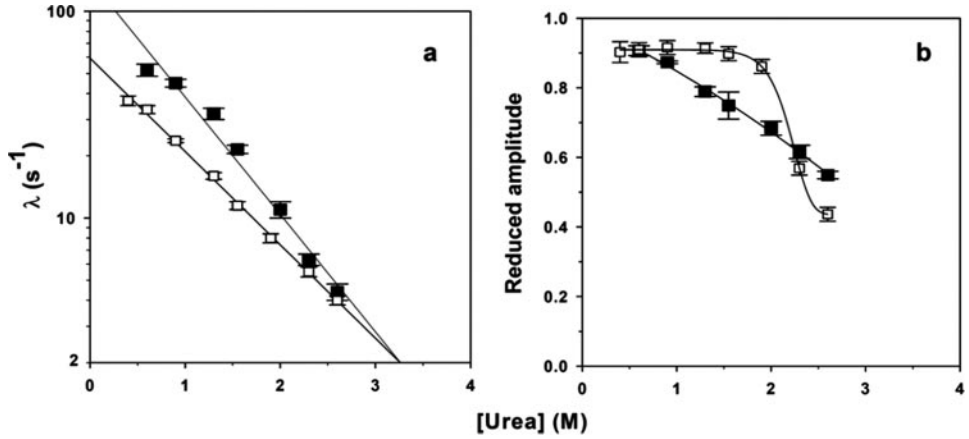

FIGURE 7. **Dependence of the refolding kinetics on urea concentration at pH 9.2.** *a*, dependence of the logarithm of the fast refolding rate constants (λ) on urea concentration as monitored by fluorescence (□) and by cysteine accessibility (■). The *solid straight lines* represent a fit of the data to the equation $\log\lambda = m[\text{urea}] + c$, where *m* and *c* are the slope and the intercept of the straight line, respectively. The values of *m* for the fluorescence-monitored and the cysteine accessibility-monitored kinetics are −0.45 and −0.57 M^{-1}, respectively. *b*, dependence of the reduced amplitudes of the fast phase of refolding on the concentration of urea as monitored by fluorescence (□) and by cysteine accessibility (■). The *continuous lines* through the data have been drawn by inspection only. In each *panel*, the *error bars* represent the S.D. values from three separate experiments.

These results not only confirm that MMTS has no direct effect on the folding process but, importantly, also show that the kinetic coupling of the folding reaction to a covalent labeling reaction does not affect the kinetics of cysteine burial; if it did, then the rate measured for cysteine burial would be expected to vary as the rate of labeling is varied 10-fold by varying the concentration of MMTS over a 10-fold range. In summary, although the rates of cysteine burial (k_f) measured by the different probes vary only over a 3-fold range, the differences are likely to be real.

For Cys3, the thiol in the unfolded protein is labeled to the same extent by 4 and 40 mM MMTS pulses (Fig. 6*a*). The extents of labeling of the thiol in N and I_L are also independent of the concentration of MMTS in the range of 4–40 mM. The extent of labeling varies from 4 to 7% in N, and from 14 to 18% in I_L, as the concentration of MMTS is increased from 4 to 40 mM. For Cys67, the extent of the labeling of the thiol in N as well as in I_L increases as the concentration of MMTS during labeling is increased from 4 to 40 mM (Fig. 6*b*). The same value is obtained for k_f, even when the extents of labeling of I_E and I_L are varied significantly by changing the strength of the labeling pulse (Fig. 6*b*), again pointing to the robustness of the cysteine burial rates that have been obtained in this study.

Dependence of the Fluorescence-monitored and Cysteine Accessibility-monitored Refolding Kinetics on the Concentration of Urea—Fig. 7*a* shows that the linear dependences on urea concentration of the logarithm of apparent rate constants of the fast phase of refolding of Cys3 are different when monitored by both of the probes. The slope of the linear dependence is more for the cysteine burial-monitored rates (see the legend to Fig. 7*a*). Fig. 7*b* shows that the dependence of the reduced amplitude (see Equations 8 and 9 for the calculation of reduced amplitudes) of the fast phase of refolding for Cys3 on urea concentration when measured by the change in fluorescence is different from the dependence of the reduced amplitude measured by the change in cysteine accessibility, although the reduced amplitudes determined using both of the probes are identical (~90%) for folding in 0.6 M urea.

DISCUSSION

The use of mass spectrometry, for the first time, for quantifying the extent of pulsed cysteine labeling at different times during folding overcomes various disadvantages of the assay used previously (22). It greatly simplifies the processing of samples after the pulsed SX experiments. Measurements are relatively easy, take very little time, and are highly reproducible;

errors in measurement of the relative amounts of labeled and unlabeled proteins have been shown to be less than 5% (Fig. 1*d*). Moreover, the experiment is designed in such a manner that the effect of the label on the stability of the protein and on its folding pathway does not have to be taken into account.

The Early Intermediate I_E Has a Loosely Compact Structure—The earliest event in the refolding of barstar, in strongly native-like conditions as in this study, is the formation of a collapsed intermediate, I_E (28, 29, 30, 31). The structure in I_E is loosely compact (Trp^{53} remains solvent-exposed in all folding conditions) and has solvent-exposed hydrophobic patches, as shown by its ability to bind a hydrophobic dye, 8-anilino-1-naphthalene-sulfonic acid (28, 29, 31). Although these previous studies yielded information about the gross structural features of I_E, not much was known about the fate of individual residues in I_E.

To determine the extent of packing interactions between different side chain residues in I_E, six mutant proteins (Cys3, Cys40, Cys62, Cys67, Cys82, and Cys89) were chosen in which the single thiol group is fully or partially buried in the native state of the protein. For these six thiol positions, the values obtained for k_E, the apparent rate constant for the labeling of the cysteine thiol in I_E, vary over only a 3-fold range, suggesting that the protection factors for labeling must have values of $\leq$3. Not surprisingly, the kinetic curves of pulse labeling of these six thiols, when extrapolated to $t = 0$ s, essentially yield, within 10%, the labeling expected for a fully exposed cysteine thiol (Fig. 4). This indicates that all segments of the polypeptide chain are exposed to solvent in I_E. This result is important, because it implies that the structure in I_E is packed loosely throughout the entire length of the polypeptide chain. The loss in entropy during a collapse leading to such a loose structure would be relatively small and would need to be compensated for by a correspondingly small enthalpy change.

Nonnative Structure Appears to Be Absent in the Early Intermediate Ensemble I_E—The pulsed SX methodology used in this study in conjunction with several single cysteine-containing mutant proteins offers an elegant way to address the question at the individual residue level as to whether any native- or nonnative-like structures form in the initial collapsed species as well as during the subsequent folding reaction. If nonnative structures are present in I_E, then it might be expected that amino acid residues that are surface residues in the folded protein might get buried transiently in such nonnative structure.

To determine whether any nonnative structure is present in I_E, four mutant proteins (Cys14, Cys25, Cys36, and Cys42) were chosen in which the thiol group is on the surface of the native protein. No transient burial of any of the four surface cysteine residues is observed in I_E, and each of the four cysteine thiols remains exposed to solvent at all times during the folding process (Fig. 4). It appears that under the conditions employed for this study, nonnative structures in which any one of these four differently located thiols gets buried transiently are not present in I_E and during the subsequent folding reaction.

The Use of Pulsed Cysteine Labeling to Study Protein Folding Kinetics—Pulsed cysteine labeling is fundamentally different from an instantaneous measurement of the extent of folding, such as fluorescence (23). In the present study, the kinetics of the major fast folding reaction, from I_E to I_L, has been studied. It is conceivable that the thiol in I_L may get labeled to a significant extent during the 4 ms labeling pulse and that such labeling of I_L at a specific site may affect the extraction of the site-specific folding rate constant. Although global unfolding of I_L to I_E is unlikely to occur to any significant extent during the strongly stabilizing conditions of the labeling pulse (26–28, 34), it is possible that I_L undergoes rapid local unfolding in the structural region where any particular cysteine thiol is located under these conditions. If the sites of the different thiols undergo different extents of local unfolding reactions, then different thiols will get labeled to different extents in I_L. If such a rapid, local unfolding reaction at the cysteine thiol location does indeed occur in I_L, and if it is rapid compared with the duration of the pulse, then because it takes about 0.5 ms for an exposed thiol to get labeled under the labeling conditions used, the cysteine thiol in I_L may have several opportunities to get labeled during the 4 ms pulse. As a result, more I_L will get labeled than expected, and the apparent rate constant, k_L, for the labeling of the cysteine thiol in I_L might be high in value. Finally, I_L has been reported to be a heterogeneous assembly of structures (32–34), and it is possible that some of these might have one or more of the cysteine thiols not buried but solvent-exposed. This would lead to significant labeling of the cysteine thiol in I_L. It was therefore important to determine whether labeling of I_L during the 4 ms pulse, through any of these routes, can affect the measurement of the kinetics of cysteine burial during folding.

Fortunately, the extent of labeling in I_L could be measured directly by applying the 4 ms labeling pulse at a time of folding (200–300 ms) when I_L is populated to the maximum extent. This allowed the value of k_L, to be determined accurately in a site-specific manner. Reliable values could also be obtained for k_E and k_L, because the value of the former is determined by the extent of labeling at the beginning of the fast folding reaction, when I_E is the predominant form, and that of the latter is determined by the extent of the labeling at the end of the fast folding reaction, when I_L is the predominant form.

Cooperativity of the Fast Folding, I_E to I_L Reaction—In this study, a 3-fold dispersion in the rates of cysteine thiol burial at different structural locations is seen during folding (Fig. 5). These rates appear to be equal to or 3-fold faster than that measured by the change in Trp^{53} fluorescence. Great care has been taken to measure the site-specific folding rates very precisely (see above). The differences observed in site-specific rates can account for why the rates of the I_E to I_L structural transition are 2–3-fold different when monitored by far UV-CD and fluorescence (29, 31). Similar or slightly larger dispersions in site-specific rates have been seen in real time NMR studies of the slow folding of other proteins (20, 41) or in pulsed HX-NMR studies of the attainment of main chain structure during fast folding (42, 43). In the real time NMR studies of ubiquitin (20) and interleukin-1β (41), unfolding was described as being highly noncooperative.

The variation of the reduced amplitude of the fast phase of refolding with urea concentration represents the equilibrium unfolding transition of I_L to I_E (26, 27). The observation of noncoincident curves when this transition was monitored using the burial of Trp^{53} and the burial of Cys^3 as probes (Fig. 7*b*) indi-

cates that the packing interactions involving Cys^3 and Trp^{53} are lost at different times during the unfolding of I_L to I_E. Hence, I_L must unfold in multiple steps, and at least one intermediate, and hence two transition states, are populated during the I_E to I_L transition. The observation that the dependence on urea concentration of the rate of the fast phase of folding (which monitors the I_E to I_L structural transition) is different when the folding reaction was monitored using the change in fluorescence of Trp^{53} and change in the accessibility of Cys^3 as probes (Fig. 7*a*) implies that the barrier height experienced by the two different parts of the polypeptide chain during folding is different, and hence, the transition state of the I_E to I_L folding reaction seen by these two residue-specific probes is different. These results, combined with the observations that slightly different apparent rate constants are measured during the fast phase of refolding by the individual residue-specific thiol probes and their noncoincidence with fluorescence-monitored kinetics for individual mutant proteins (Fig. 5), suggest that the I_E to I_L structural transition might occur in multiple steps at the level of individual side chains.

Lattice model simulations suggest a wide distribution of side chain relaxation rates during protein folding (44), but experimental kinetic evidence has been lacking so far (45). Such evidence is necessary to understand whether packing interactions in the protein core, which are so important in determining native state and transition state stability (46–49), indeed form simultaneously or whether they form via a slow diffusive search by the polypeptide chain. The use of 10 residue-specific probes located in different parts of the protein structure made it possible to measure the distribution of rate constants of side chain burial during the fast refolding reaction of barstar. It is observed that Cys^3, Cys^{62}, and Cys^{67} get buried at a rate slower than that of Cys^{40}, Cys^{82}, and Cys^{89} (Fig. 5, *a* and *b*). Cys^3 and Cys^{67} are in the same hydrophobic pocket, and not surprisingly, the relative rates of burial of their cysteine thiols are also the same (Fig. 5*b*). Cys^{82} makes one residue loop just at the beginning of the third β-strand, and Cys^{89} is the last residue of this β-strand. It is observed that the relative rates of burial of their cysteine thiols are similar and faster than those of the other mutant proteins (Fig. 5*b*). It appears that the third β-strand acquires protective structure earlier than the rest of the protein. These results suggest that the packing interactions that are required to stabilize the native protein are acquired by different regions of the protein structure at different times during folding.

It is possible that the different residue-specific thiol probes monitor different sequential steps during the I_E to I_L folding reaction. The fastest rate for the I_E to I_L folding reaction seen by any residue-specific probe is 120 s^{-1} (for Cys89; see Table 1). The rate of burial of Trp^{53}, which is the same as the rate of burial of thiol of Cys62, is slowest among all of the residue-specific probes (~34 s^{-1}). The failure to observe the expected 1–2 ms lag phase in the formation of I_L, when monitored either by the change in fluorescence or by any other residue-specific probe (Figs. 3 and 4) cannot, however, be used to rule out a sequential series of folding events, because the dead times of the kinetic measurements were themselves about 2 ms. Lag phases in protein folding are notoriously difficult to detect, in part because, given the complex nature of the folding process, it is very unlikely that the change in signal associated predominantly with the slowest step does not also occur to at least some extent in the earlier steps. Alternatively, the differences observed in the site-specific rates could mean that folding is not synchronized across different regions of the protein molecule and that the multiple steps occur in parallel, independently of each other. Given the smallness of the dispersion observed in the site-specific rates, it is difficult to resolve the origin of the dispersion.

Heterogeneity of Structure in I_L—Time-resolved FRET studies have shown that the structure of I_L is heterogeneous and that the extent of heterogeneity decreases as folding conditions become more native-like (34). This result was indicated by the observation that more and more intramolecular distances in I_L become native-like as folding conditions are made more native-like. In this study, folding was carried out in strongly stabilizing conditions (0.6 M urea), and it is seen that the structure of I_L is quite native-like; cysteine thiols that are on the surface in N are on the surface in I_L, and cysteine thiols that are buried in N are also all buried in I_L.

It is instructive to examine why the extent of labeling of the cysteine thiol in I_L is seen to be more than in the native protein for Cys3, Cys40, and Cys82 (Fig. 4). The trivial explanation for this observation, as exemplified in the analysis of the data using Equation 6, is that the protection factor in I_L is less than that in the native protein at each of these cysteine locations. However, at least in the case of Cys3, this does not appear to be true. The extent of labeling of the thiol of Cys3 in I_L remains at 16 ± 2% even when the strength of the pulse is increased 10-fold by increasing the concentration of MMTS in the labeling pulse from 4 to 40 mM (Fig. 6*a*). Similarly, the extent of labeling of the thiol of Cys3 in the native protein remains the same at 5 ± 2% (Fig. 6*a*). Hence, the protection factor of the thiol of Cys3 should be similar in I_L and in the native protein. Why then is the extent of labeling of the thiol in I_L more than that in the native protein? A plausible explanation is that all of the folding molecules have not formed I_L at the time when the formation of I_L is complete (300 ms) (*i.e.* when I_L is populated to the maximum extent). In a simple sequential model, each probe should become protected in only a single step of folding, once it is established (as done here through a 10-fold variation in the strength of the labeling pulse) that the probe (the cysteine thiol) has only a marginal protection factor in the conditions of the labeling pulse (42, 50).

One explanation for the observation that all folding molecules have not formed I_L when the I_E to I_L reaction is complete (as it is at 300 ms) is that a certain fraction (~15%) of the molecules is protected by another folding step that necessarily must be on another folding pathway. Pulsed HX studies on several other proteins (43, 51, 52) have also shown that the folding kinetics for these proteins are not synchronized at the level of individual residues, and parallel routes for folding might be operating. There is, however, an alternative explanation that does not invoke multiple pathways. In an early pulsed HX study of the folding of ribonuclease A (42) and subsequently in similar studies of other proteins (53), it was observed that a certain fraction of the molecules did not form an early folding intermediate when the reaction leading to its formation was complete.

It was proposed that there was a transient kinetic barrier that prevented some of the unfolded molecules from folding, and it was suggested that an intermediate conformation with nonnative structure might function as such a kinetic trap (42). More recently, a similar explanation has been proposed for describing the folding kinetics of hen lysozyme in lieu of the multiple pathway explanation (54). It remains to be seen if such kinetic traps occur on the folding pathway of barstar.

Implications for φ-Value Analysis—The fates of individual residues during the refolding of cold-denatured barstar have been characterized previously using the φ-value analysis methodology (55), and four of the 10 residue-specific probes (Ser^{14}, Ala^{25}, Ala^{36}, and Ala^{67}) used in this study have been used as reporter residues for the φ-value measurements. The φ-values reported for these four residues in the initial intermediate were ~0.5 for Ser^{14}, ~0.2 for Ala^{25}, ~0.2 for Ala^{36}, and ~0.1 for Ala^{67}. Based on this observation, it was concluded that interactions involving these residues are partially formed in the initial folding intermediate, I_E (55). However, the meaning of such partial φ-values, which have also been observed for most of the proteins studied to date (56), remains uncertain (57), because they can also arise if the initial intermediate is an ensemble of multiple structural forms, which are presumably formed on parallel pathways (58). In the present study, there is no indication of structure formation at the four probe sites (residue positions 14, 25, 36, and 67) in I_E, but there is also no direct evidence for multiple folding routes. It is possible, therefore, that the partial φ-values reflect effects of mutations on residual structures in the unfolded state. Although the unfolded state of a protein is usually assumed to be a random coil and, hence, not to be affected by mutation, this assumption may not be valid (59–61). Residual structures (both native- and nonnative-like) are found to exist in the unfolded states of many proteins (62, 63), including the cold-denatured state of barstar (64, 65), which was used as the starting state for the φ-value analysis study of the folding of barstar (55). It has been shown that such residual structures can be modulated by a change in solvent conditions and by mutagenesis and that such modulations affect the stability of the unfolded protein (60, 66). Hence, there is ambiguity in the structural interpretation of the φ-value analysis of the folding of barstar.

Understanding the cooperativity of protein folding reactions is important for delineating the partial unfolding reactions that lead to the formation of aggregation-competent intermediate structures (67, 68), which lead to the amyloid fibril formation characteristic of many neurodegenerative diseases (69, 70). From the perspective of understanding the roles of intermediates in protein folding, especially of the high energy intermediates thought to be populated during the folding of proteins that appear otherwise to be two-state folders (71–74), it is important to develop methods, such as the one reported in this study, that yield site-specific information on the cooperativity of the major fast folding reaction of any protein. From a practical viewpoint, if the major fast folding reactions of the apparent two-state folding proteins will also similarly appear to be multistep reactions, when characterized by multiple site-specific probes, it would have a major implication on the study of such reactions by the elegant φ-value analysis, where typically only one gross structural probe is used to measure folding kinetics (75).

REFERENCES

1. Dill, K. A., Fiebig, K. M., and Chan, H. S. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 1942–1946
2. Klimov, D. K., and Thirumalai, D. (1998) *Fold. Des.* **3,** 127–139
3. Chan, H. S., Shimizu, S., and Kaya, H. (2004) *Methods Enzymol.* **380,** 350–379
4. Reich, L., and Weikl, T. R. (2006) *Proteins* **63,** 1052–1058
5. Finkelstein, A. V., and Ptitsyn, O. B. (2002) *Protein Physics*, pp. 205–277, Academic Press, London
6. Swaminathan, R., Nath, U., Udgaonkar, J. B., Periasamy, N., and Krishnamoorthy, G. (1996) *Biochemistry* **35,** 9150–9157
7. Holtzer, M. E., Lovett, E. G., D'Avignon, D. A., and Holtzer, A. (1997) *Biophys. J.* **73,** 1031–1041
8. Song, J., Jamin, N., Gilquin, B., Vita, C., and Menez, A. (1999) *Nat. Struct. Biol.* **6,** 129–134
9. Lakshmikanth, G. S., Sridevi, K., Krishnamoorthy, G., and Udgaonkar, J. B. (2001) *Nat. Struct. Biol.* **8,** 799–804
10. Ahmed, Z., Beta, I. A., Mikhonin, A. V., and Asher, S. A. (2005) *J. Am. Chem. Soc.* **127,** 10943–10950
11. Sadqi, M., Fushman, D., and Munoz, V. (2006) *Nature* **442,** 317–321
12. Li, H., and Frieden, C. (2007) *Biochemistry* **46,** 4337–4347
13. Eaton, W. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 5897–5899
14. Huang, F., Sato, S., Sharpe, T. D., Ying, L., and Fersht, A. R. (2007) *Proc. Natl. Acad. Sci. U. S. A.* **104,** 123–127
15. Sabelko, J., Ervin, J., and Gruebele, M. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6031–6036
16. Osvath, S., Sabelko, J. J., and Gruebele, M. (2003) *J. Mol. Biol.* **333,** 187–199
17. Ma, H., and Gruebele, M. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2283–2287
18. Sinha, K. K., and Udgaonkar, J. B. (2005) *J. Mol. Biol.* **353,** 704–718
19. Sinha, K. K., and Udgaonkar, J. B. (2007) *J. Mol. Biol.* **370,** 385–405
20. Schanda, P., Forge, V., and Brutscher, B. (2007) *Proc. Natl. Acad. Sci. U. S. A.* **104,** 11257–11262
21. Ballery, N., Desmadril, M., Minard, P., and Yon, J. M. (1993) *Biochemistry* **32,** 708–714
22. Ha, J. H., and Loh, S. N. (1998) *Nat. Struct. Biol.* **5,** 730–737
23. Ramachandran, S., Rami, B. R., and Udgaonkar, J. B. (2000) *J. Mol. Biol.* **297,** 733–745
24. Sridevi, K., and Udgaonkar, J. B. (2002) *Biochemistry* **41,** 1568–1578
25. Silverman, J. A., and Harbury, P. B. (2002) *J. Biol. Chem.* **277,** 30968–30975
26. Schreiber, G., and Fersht, A. R. (1993) *Biochemistry* **32,** 11195–11203
27. Shastry, M. C., Agashe, V. R., and Udgaonkar, J. B. (1994) *Protein Sci.* **3,** 1409–1417
28. Shastry, M. C., and Udgaonkar, J. B. (1995) *J. Mol. Biol.* **247,** 1013–1027
29. Agashe, V. R., Shastry, M. C., and Udgaonkar, J. B. (1995) *Nature* **377,** 754–757
30. Pradeep, L., and Udgaonkar, J. B. (2002) *J. Mol. Biol.* **324,** 331–347
31. Pradeep, L., and Udgaonkar, J. B. (2004) *J. Biol. Chem.* **279,** 40303–40313
32. Bhuyan, A. K., and Udgaonkar, J. B. (1999) *Biochemistry* **38,** 9158–9168
33. Sridevi, K., Juneja, J., Bhuyan, A. K., Krishnamoorthy, G., and Udgaonkar, J. B. (2000) *J. Mol. Biol.* **302,** 479–495
34. Sridevi, K., Lakshmikanth, G. S., Krishnamoorthy, G., and Udgaonkar, J. B. (2004) *J. Mol. Biol.* **337,** 699–711
35. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA
36. Khurana, R., Hate, A. T., Nath, U., and Udgaonkar, J. B. (1995) *Protein Sci.* **4,** 1133–1144
37. Roberts, D. D., Lewis, S. D., Ballou, D. P., Olson, S. T., and Shafer, J. A. (1986) *Biochemistry* **25,** 5595–5601
38. Agashe, V. R., and Udgaonkar, J. B. (1995) *Biochemistry* **34,** 3286–3299
39. Sridevi, K. (2003) *Structural and Temporal Characterization of the Folding Pathway of Barstar.* Ph.D. thesis, pp. 62–82
40. Rami, B. R., and Udgaonkar, J. B. (2001) *Biochemistry* **40,** 15267–15279

41. Roy, M., and Jennings, P. A. (2003) *J. Mol. Biol.* **328,** 693–703
42. Udgaonkar, J. B., and Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8197–8201
43. Jacobs, M. D., and Fox, R. O. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 449–453
44. Kussell, E., Shimada, J., and Shakhnovich, E. I. (2003) *Proteins* **52,** 303–321
45. Frieden, C. (2003) *Biochemistry* **42,** 12439–12446
46. Lim, W. A., and Sauer, R. T. (1989) *Nature* **339,** 31–36
47. Levitt, M., Gerstein, M., Huang, E., Subbiah, S., and Tsai, J. (1997) *Annu. Rev. Biochem.* **66,** 549–579
48. Northey, J. G., Di Nardo, A. A., and Davidson, A. R. (2002) *Nat. Struct. Biol.* **9,** 126–130
49. Rami, B. R., Krishnamoorthy, G., and Udgaonkar, J. B. (2003) *Biochemistry* **42,** 7986–8000
50. Roder, H., Elove, G. A., and Englander, S. W. (1998) *Nature* **335,** 700–704
51. Radford, S. E., Dobson, C. M., and Evans, P. A. (1992) *Nature* **358,** 302–307
52. Baldwin, R. L. (1995) *J. Biomol. NMR* **5,** 103–109
53. Englander, S. W., and Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct.* **21,** 243–265
54. Krishna, M. M., and Englander, S. W. (2007) *Protein Sci.* **16,** 449–464
55. Nolting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 826–830
56. Goldenberg, D. P. (1999) *Nat. Struct. Biol.* **6,** 987–990
57. Ozkan, S. B., Bahar, I., and Dill, K. A. (2001) *Nat. Struct. Biol.* **8,** 765–769
58. Weikl, T. R., and Dill, K. A. (2007) *J. Mol. Biol.* **365,** 1578–1586
59. Cho, J.-A., Sato, S., and Raleigh, D. P. (2004) *J. Mol. Biol.* **338,** 827–837
60. Pradeep, L., and Udgaonkar, J. B. (2004) *Biochemistry* **43,** 11393–11402
61. Cho, J.-A., and Raleigh, D. P. (2006) *J. Mol. Biol.* **359,** 1437–1446
62. Shortle, D. (2002) *Adv. Protein. Chem.* **62,** 1–23
63. Juneja, J., and Udgaonkar, J. B. (2003) *Curr. Sci.* **84,** 157–172
64. Wong, K. B., Freund, S. M., and Fersht, A. R. (1996) *J. Mol. Biol.* **259,** 805–818
65. Nolting, B., Golbik, R., Soler-Gonzalez, A. S., and Fersht, A. R. (1997) *Biochemistry* **36,** 9899–9905
66. Anil, B., Li, Y., Cho, J. H., and Raleigh, D. P. (2006) *Biochemistry* **45,** 10110–10116
67. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., and Pepys, M. B. (1997) *Nature* **385,** 787–793
68. Hamid Wani, A., and Udgaonkar, J. B. (2006) *Biochemistry* **45,** 11226–11238
69. Fink, A. L. (1998) *Fold. Des.* **3,** 9–23
70. Chiti, F., and Dobson, C. M. (2006) *Annu. Rev. Biochem.* **75,** 333–366
71. Bowie, J. U., and Sauer, R. T. (1989) *Biochemistry* **28,** 7139–7143
72. Zaidi, F. N., Nath, U., and Udgaonkar, J. B. (1997) *Nat. Struct. Biol.* **4,** 1016–1024
73. Juneja, J., and Udgaonkar, J. B. (2002) *Biochemistry* **41,** 2641–2654
74. Sanchez, I. E., and Kiefhaber, T. (2003) *J. Mol. Biol.* **325,** 367–376
75. Fersht, A. R., and Sato, S. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7976–7981