doi:10.1016/j.jmb.2007.04.061 J. Mol. Biol. (2007) 370, 385–405

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Dissecting the Non-specific and Specific Components of the Initial Folding Reaction of Barstar by Multi-site FRET Measurements

Kalyan K. Sinha and Jayant B. Udgaonkar*

National Centre for Biological Sciences, Tata Institute of Fundamental Research UAS-GKVK Campus Bangalore 560065, India

*Corresponding author

Initial polypeptide chain collapse plays a major role in the development of subsequent structure during protein folding, but it has been difficult to elucidate the coupling between its cooperativity and specificity. To better understand this important aspect of protein folding, nine different intramolecular distances in the protein have been measured by fluorescence resonance energy transfer (FRET) in the product(s) of the initial, sub-millisecond collapse reaction during the folding of barstar, under different folding conditions. All nine distances contract in these initial folding products, when the denaturant concentration is reduced. Two of these distances were also measured in peptides corresponding to sequence segments 38–55 and 51–69 of the protein. Surprisingly, both distances do not contract in the peptides which remain fully unfolded when the denaturant concentration is reduced. This suggests that the contraction of at least some segments of the polypeptide chain may be facilitated only by contraction of other segments. In the case of the initial product of folding of the protein, the dependence on denaturant concentration of the relative change in each distance suggests that there are two components to the initial folding reaction. One is a nonspecific component, which appears to be driven by the change in denaturant concentration that is used to initiate refolding. This component corresponds to the collapse of completely unfolded protein (U) to unfolded protein in refolding conditions (U_C). The extent of nonspecific collapse can be predicted by the response of completely unfolded protein to a change in denaturant concentration. All distances undergo such solvent-induced contraction, but each distance contracts to a different extent. There is also a specific component to initial sub-millisecond folding, in which some distances (but not all) contract more than that predicted by solvent-induced contraction. The observation that only some of the distances undergo contraction over and above solvent-induced contraction, suggest that this specific component is associated with the formation of a specific intermediate (I_E). FRET efficiency and distance change differently for the different donor–acceptor pairs, with a change in denaturant concentration, indicating that the formation or dissolution of structure in U_C and I_E does not happen in a synchronized manner across different regions of the protein molecule. Also, all nine FRET efficiencies and intramolecular distances in the product(s) of sub-ms folding, change continuously with a change in denaturant concentration. Hence, it appears that the transitions from U to U_C and to I_E are gradual transformations, and not all-or-none structural transitions. Nevertheless, the product of these gradual transitions, I_E, possesses specific structure.

Keywords: protein folding; polypeptide chain collapse; cooperativity; specificity

Abbreviations used: FRET, fluorescence resonance energy transfer; TNB, thionitrobenzoic acid; D-A, donor-acceptor.

E-mail address of the corresponding author: jayant@ncbs.res.in

0022-2836/$ - see front matter © 2007 Published by Elsevier Ltd.

Introduction

Unfolded protein chains in high concentrations of denaturant are thought to behave as random coils,[1–4] even when they might possess non-random structure,[5] and they contract when placed in native-like conditions. Natively folded proteins that have been destabilized by mutation so strongly that they no longer fold under native-like conditions,[6–8] as well as natively unfolded proteins,[9–11] expand when put into high concentrations of denaturant. In this respect, the behavior of protein chains resembles that of homopolymer chains transferred from a bad to a good solvent.[12] Since protein folding reactions are initiated typically by dilution of denaturant, the first step in a protein folding reaction is expected to be a solvent-induced collapse of the unfolded state, which is thought to occur within 100 μs of the initiation of folding,[13,14] and which is likely to be non-specific by its very nature.[15] Not surprisingly, there has been some effort aimed at showing that the product of the sub-millisecond folding reaction of a protein could just be the unfolded state in refolding conditions, and not a structured intermediate.[15–18] On the other hand, there are some proteins that undergo polypeptide chain collapse and refold completely in the sub-millisecond time domain.[19,20]

In earlier kinetic studies with millisecond resolution, it was usually not possible to resolve temporally chain collapse from specific structure formation. The products of the sub-millisecond folding reactions of many proteins appeared to be compact and to possess secondary structure, suggesting that collapse and specific structure formation occurred simultaneously.[21,22] In fact, many models of protein folding have suggested that secondary structures formed before chain collapse occurred.[23–25] One of the first indications that collapse could precede secondary structure formation came from a study of the folding of barstar,[26] where it was shown that the product of the sub-millisecond folding reaction under marginally stabilizing conditions was collapsed but did not possess secondary structure. Only in strongly stabilizing folding conditions did the collapsed product possess secondary structure, but secondary structure formation and collapse could not be resolved temporally.[27] Now it appears that polypeptide chain collapse may precede structure formation even for the so-called two-state folding proteins, for which all structure formation is supposedly synchronized.[28] It seems possible for polypeptide chain collapse to follow secondary structure formation only when folding is carried out at cryogenic temperatures, where hydrophobic interactions would be greatly weakened.[29] Structureless protein globules have been identified under different experimental[26,29–34] and theoretical[35,36] contexts in protein folding reactions. The advent of sub-millisecond kinetic methods has made it possible to study the actual process of collapse in a few cases. A non-specific collapse of the polypeptide chain has been shown to precede structure formation during the folding of ribonuclease A.[37] In the case of cytochrome *c*[38] and monellin,[39] no significant secondary structure formation is seen to accompany the fastest observable collapse reaction, indicating that collapse precedes secondary structure formation. In the case of apomyogobin, it was not possible to resolve temporally chain collapse and secondary structure formation, in mixing experiments in the 100 μs time domain,[40] although fast structural events have been detected in the 10 μs time domain using other methods.[41]

Specificity of polypeptide chain collapse has been inferred indirectly from two different sets of experimental observations. Firstly, when the products of sub-millisecond folding reactions of several proteins were examined at a few milliseconds of folding, their unfolding appeared to have sigmoidal dependence on the concentration of urea.[22,42,43] A sigmoidal dependence is commonly interpreted as arising from a cooperative transition. But it has been shown that a sigmoidal dependence does not necessarily have to arise from a cooperative transition.[44,45] In fact, for several proteins, different structural probes yield either non-overlapping sigmoidal transitions[27,46] or non-sigmoidal transitions[47] for the unfolding of the products of initial sub-millisecond folding, indicating that they have formed in a stepwise manner from the unfolded protein in refolding conditions. Direct sub-millisecond measurements of folding also indicate that the initial collapse reaction is followed quickly by a second structure-forming reaction, so that the product observed at a few milliseconds of folding may be the consequence of more than one folding transition.[37–39] Thus, the product of folding examined at a few milliseconds of folding may be the result of multiple folding transitions that have occurred consecutively and/or concurrently.

Secondly, in folding experiments initiated by sub-millisecond mixing, the observation of very fast exponential kinetics for the folding of cytochrome *c* and other proteins[48] was interpreted to represent a barrier-limited, all-or-none folding transition. It was held that such a cooperative folding transition could only be specific, presumably because the formation of specific interactions must be slowed by a substantial barrier.[49] But the more recent observation that large unstructured fragments of cytochrome *c*,[50] can collapse with exponential kinetics nearly identical with the folding kinetics of collapse of intact cytochrome *c*, has indicated that exponential kinetics are not necessarily indicative of specific structure formation.[51,52] It seems that a polypeptide chain collapse reaction unaccompanied by the formation of any specific structure, can also display exponential kinetics.

The identification of exponential collapse kinetics with specific structure formation has also arisen because it is widely believed that if a folding transition occurs in a gradual manner, for example by many diffusive transitions over small energy barriers, then the observed kinetics cannot be

exponential. Consequently, the observation of non-exponential kinetics during folding has been taken to mean that the folding is gradual or downhill in nature.[53,54] Non-exponential kinetics have also been attributed to parallel downhill and activated folding pathways.[55] A diffusive folding reaction might itself display exponential kinetics.[45,52] It appears that just as non-exponential kinetics need not necessarily be simply taken as a signature of a gradual process over many small barriers, exponential kinetics need not necessarily signify a barrier-limited, all-or-none process. The initial collapse process, like the overall folding process, may be a combination of first-order as well as higher-order transitions,[56] and the kinetics for such processes remain ill-defined.

There is an additional aspect to polypeptide chain collapse during folding that adds further confusion to the relationship between cooperativity and specificity. This is the conformational heterogeneity that must be present in the first step in protein folding. Every unfolded molecule in a high concentration of urea is expected to exist in a different conformation, and each conformation is expected to change every 100 ps or so. In addition, large-scale diffusive motions occur in the 10 μs time domain.[57] To better characterize the specificity and cooperativity of a polypeptide chain collapse reaction, as well as the heterogeneity inherent in the collapse, it becomes imperative to use multiple structural probes that can report on structure formation in different parts of the protein molecule as it folds. In this study, such an approach has been used to characterize the product of the sub-millisecond folding reactions of barstar.

Barstar is an 89 residue protein that serves as intracellular inhibitor of the extracellular protease barnase in *Bacillus amyloliquefaciens*. Its folding pathway has been characterized, and several intermediates are known to populate its folding pathway.[58,59] Under strongly stabilizing conditions, an early intermediate, I_E, is known to be populated on the folding pathway during the initial few milliseconds of the folding.[27] It was shown earlier that I_E is an ensemble of at least three structural forms, and that different forms populate the folding pathway under different folding conditions.[27,60] In marginally stabilizing conditions, the product of the initial folding phase has very little structure or does not accumulate to a detectable extent, but retains the ability to bind the hydrophobic dye 8-anilino-1-naphthalene sulfonic acid (ANS): it therefore appears to be a collapsed structure (U_C).[26] The formation of U_C and I_E appears to be a non-cooperative process, as suggested by the non-coincidence of the burst phase amplitudes observed in fluorescence and CD-monitored kinetic studies with millisecond time resolution.[27]

In this study, the objective was to obtain an insight into how cooperativity and specificity are coupled during the initial folding reaction, including polypeptide chain collapse. Nine intramolecular distances in different structural segments of barstar were measured in the product of the initial folding reaction, as a function of denaturant concentration, at a few milliseconds of folding. This was done by measurements of the fluorescence resonance energy transfer (FRET)-monitored refolding kinetics using different single cysteine-containing, single tryptophan-containing variants of barstar, each having the acceptor moiety in a different structural region of the protein (Figure 1). The proteins used are Cys3, Cys14, Cys36, Cys40, Cys42, Cys67, Cys79, Cys82 and Cys89. The single tryptophan (Trp53) serves as the FRET donor, and a small thionitrobenzoic acid (TNB)-adduct coupled to the thiol group of the single cysteine residue serves as the FRET acceptor. The utility of multi-site FRET in characterizing protein folding and unfolding reactions has been demonstrated.[61–64] In this study, the use of multiple FRET pairs has allowed an analysis of the non-cooperative behavior of the initial folding transition, including collapse, and has also given useful clues regarding its specificity. It appears from the present study that the product of the initial folding reaction is a specifically collapsed species at a few milliseconds of folding. Yet, it appears to form by a gradual, not an all-or-none, transition.

Results

Refolding kinetics of unlabeled proteins

Figure 2(a) and (b) show the first 0.5 s of representative kinetic traces of refolding in 0.6 M urea for two of the nine mutant proteins, Cys14 and Cys79, respectively. The refolding kinetics of the nine different single Cys-containing, single Trp-containing mutant proteins were monitored at 320 nm by measuring the increase in fluorescence intensity of the Trp53 upon refolding. There is no burst phase (<6 ms) change in fluorescence during the refolding of any of the proteins studied: the fit to the kinetic trace extrapolates back to the fluorescence signal of the unfolded protein. In other words, the observed kinetic amplitude accounts for the entire equilibrium amplitude expected from the equilibrium unfolding transition of the protein.

Refolding kinetics of the TNB-labeled proteins

Figure 2(c) and (d) show the initial 0.5 s of the kinetics of refolding in 0.6 M urea of Cys14-TNB and Cys79-TNB, respectively. The refolding kinetics of the nine different TNB-labeled proteins were monitored using the decrease in fluorescence intensity at 380 nm. The rationale for selecting this wavelength for monitoring the kinetics of the folding of the labeled proteins has been explained elsewhere.[66] In all the cases, there is a significant and very fast unobservable decrease in signal amplitude within the dead-time (~6 ms) of the stopped-flow mixing device. This very fast burst phase change, which precedes the observed fast refolding phase, must occur in the sub-millisecond time domain.

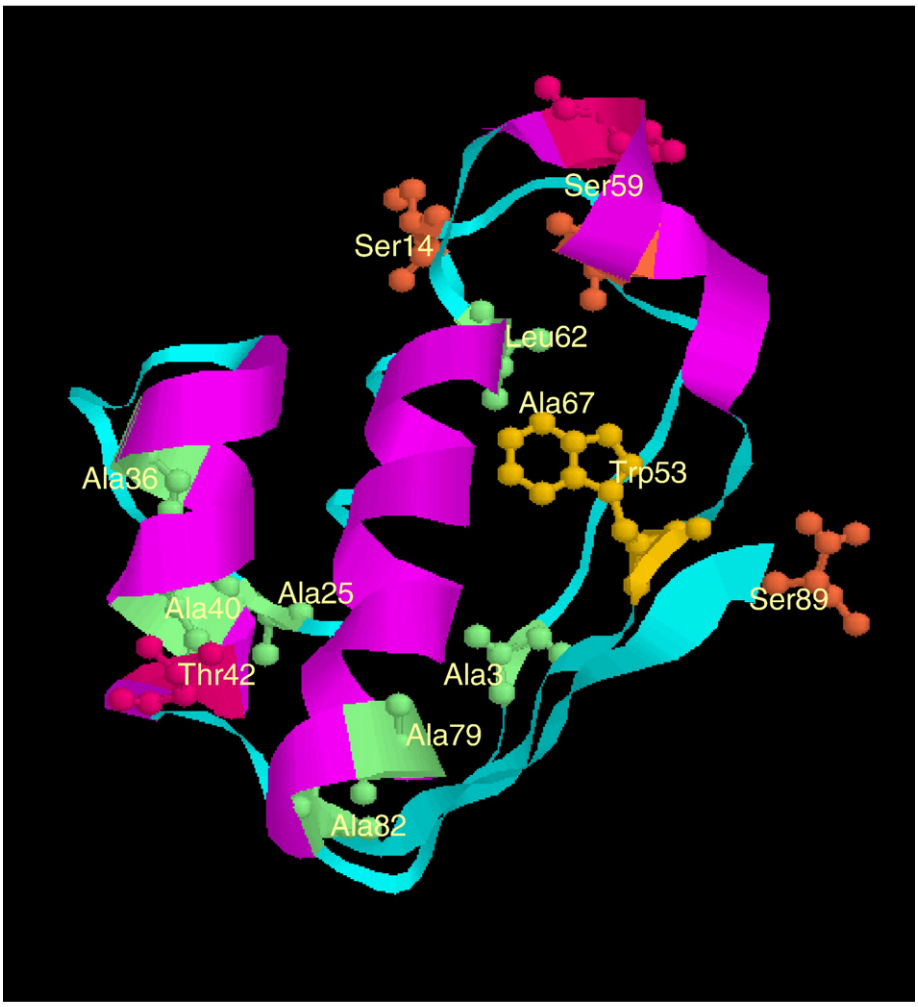

Figure 1. Structure of barstar. The locations of the different residues, which were independently mutated to cysteine to generate the nine different single Cys-containing mutant proteins, are shown. Each of these mutant proteins has a single tryptophan residue, Trp53, which is located in the core of the protein. Cys14, Cys42 and Cys79 have replaced residues that were solvent-exposed. Cys3 and Cys67 are in the main hydrophobic core of the protein. Cys40 and Cys82 form part of a separate hydrophobic pocket. The structure was generated from the PDB file 1btb using Rasmol.[107]

Figure 3(a)–(i) compare the observed fast rate constants of folding of the different labeled proteins to those of the corresponding unlabeled proteins. The apparent rate constants appear to be similar for the unlabeled proteins and the corresponding labeled proteins. The similarities in the kinetics of the refolding of the unlabeled and the TNB-labeled proteins allow the results with the different mutant proteins, labeled as well as unlabeled, to be compared directly to one another.

Dependence of the burst phase amplitude on urea concentration

Figure 4(a)–(i) compare the kinetic amplitudes of the refolding reactions monitored at 380 nm, to the equilibrium amplitudes at 380 nm, for both the unlabeled and the TNB-labeled proteins. For all the unlabeled proteins, the equilibrium unfolding transition shows a gradual increase in fluorescence intensity with increasing concentration of urea, without any sigmoidal dependence. The total increase is small and amounts to ~30–35% of the fluorescence signal at low concentrations of urea. The $t=0$ points obtained from the extrapolation of the kinetic traces of refolding measured in different concentrations of urea, also show a similar gradual increase in fluorescence intensity with increasing concentration of urea. In the case of Cys89, the kinetic $t=0$ points are coincident with the equilibrium unfolding transition, whereas for Cys3, Cys14, Cys36, Cys40, Cys42, Cys67, Cys79 and Cys82, the $t=0$ points show small deviations from the equilibrium unfolding transitions at lower concentrations of urea. It is notable that this dependence of the $t=0$ fluorescence intensity on the concentration of urea is very similar to the dependence of the fluorescence intensity of NATA (*N*-acetyl-L-tryptophanamide) on the concentration of urea (data not shown).

In the case of the TNB-labeled proteins, the equilibrium unfolding transition is sigmoidal in all

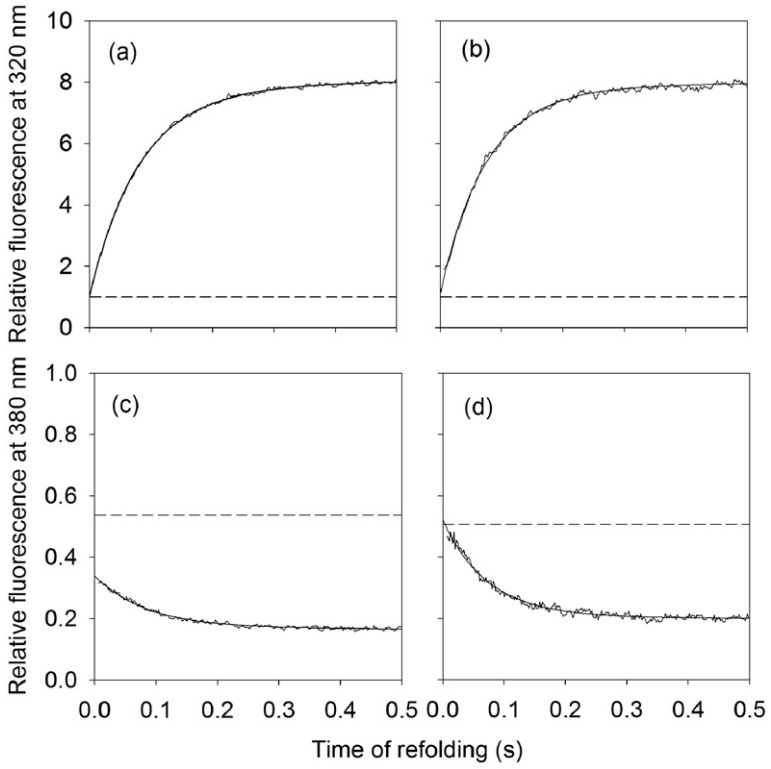

Figure 2. Refolding kinetics of unlabeled and TNB-labeled proteins at 25 °C. Shown here are representative kinetic traces of refolding in 0.6 M urea for two proteins, Cys14 and Cys79. The top panels show the refolding traces for the unlabeled proteins (a) Cys14 and (b) Cys79, monitored at 320 nm. The lower panels show the refolding traces for the two TNB-labeled proteins (c) Cys14-TNB and (d) Cys79-TNB, monitored at 380 nm. All the traces have been normalized to a value of 1 for the unfolded protein signal in 8.3 M urea. The continuous lines through the data are fits to a three-exponential equation. In each panel, the broken line represents the unfolded protein signal in refolding conditions, which was obtained from linear extrapolation of the unfolded protein baseline of the equilibrium unfolding transition to 0.6 M urea. In the case of Cys14-TNB, the fit to the refolding data does not extrapolate back to the signal of the unfolded protein in 0.6 M urea. In the case of Cys79-TNB, the fit to the kinetic curve extrapolates back at $t=0$ to the signal of the unfolded protein in 0.6 M urea.

cases. For Cys3-TNB, where the labeling seems to have affected the stability (Figure 4(a)), very little of the native protein baseline can be observed. In kinetic studies with all of the proteins, a significant amount of the fluorescence signal decreases in an unobservable sub-millisecond phase, when the folding was commenced from the 8 M urea unfolded form by denaturant dilution to lower concentrations of urea. This initial missing amplitude increases significantly with decreasing concentration of urea present during refolding. In 0.6 M urea, the missing amplitudes are 61%, 66%, 55%, 63%, 64%, 66%, 56%, 48% and 48% of the total amplitude of the folding transition from the unfolded protein in 8 M urea to the native protein in 0.6 M urea, for Cys3-TNB, Cys14-TNB, Cys36-TNB, Cys40-TNB, Cys42-TNB, Cys67-TNB, Cys79-TNB, Cys82-TNB and Cys89-TNB, respectively.

Interestingly, the dependence of the missing burst phase amplitude on the concentration of urea is different for the different labeled proteins. In the case of Cys79-TNB and Cys82-TNB (Figure 4(g) and (h)), the dependence of the extrapolated $t=0$ points on the concentration of urea are colinear with that of the fluorescence signal in the unfolded baseline region of the equilibrium unfolding transition. In other words, the $t=0$ points fall on the linearly extrapolated unfolded protein baseline of the equilibrium unfolding curve. It should be noted that two other mutant labeled proteins studied earlier,[57] Cys25-TNB and Cys62-TNB, show similar dependences of the $t=0$ points on the concentration of urea. For these proteins, for which the dependence of the $t=0$ points on the concentration of urea is colinear with the dependence of the unfolded protein baseline, it appears that the unfolded polypeptide chain responds non-specifically to a change in solvent conditions.

In the case of Cys3-TNB (Figure 4(a)), Cys14-TNB (Figure 4(b)), Cys36-TNB (Figure 4(c)), Cys40-TNB (Figure 4(d)), Cys67-TNB (Figure 4(f)), and Cys89-TNB (Figure 4(i)), the $t=0$ points show, however, a significant deviation from the linearly extrapolated unfolded protein baseline at lower concentrations of urea. In the case of Cys42-TNB (Figure 4(e)), the deviation is relatively small. For these proteins, the missing amplitude is larger than expected from linear extrapolation of the unfolded protein baseline. At any particular low concentration of urea, the difference between the fluorescence signal obtained by linear extrapolation of the unfolded protein baseline, and the fluorescence signal obtained by extrapolation to $t=0$ of the observed kinetic curve for folding in the same concentration of urea, is taken to represent the specific component of the burst phase amplitude. This specific component accounts for 19% (Cys3-TNB), 23% (Cys14-TNB), 13% (Cys36-TNB), 17% (Cys40-TNB), 9% (Cys42-TNB), 14% (Cys67-TNB), and 17% (Cys89-TNB) of the total amplitude of the transition from the unfolded protein in 8 M urea to the native state in 0.6 M urea. If it is assumed that the fluorescence signal obtained by linear extrapolation to a particular low concentration of urea of the

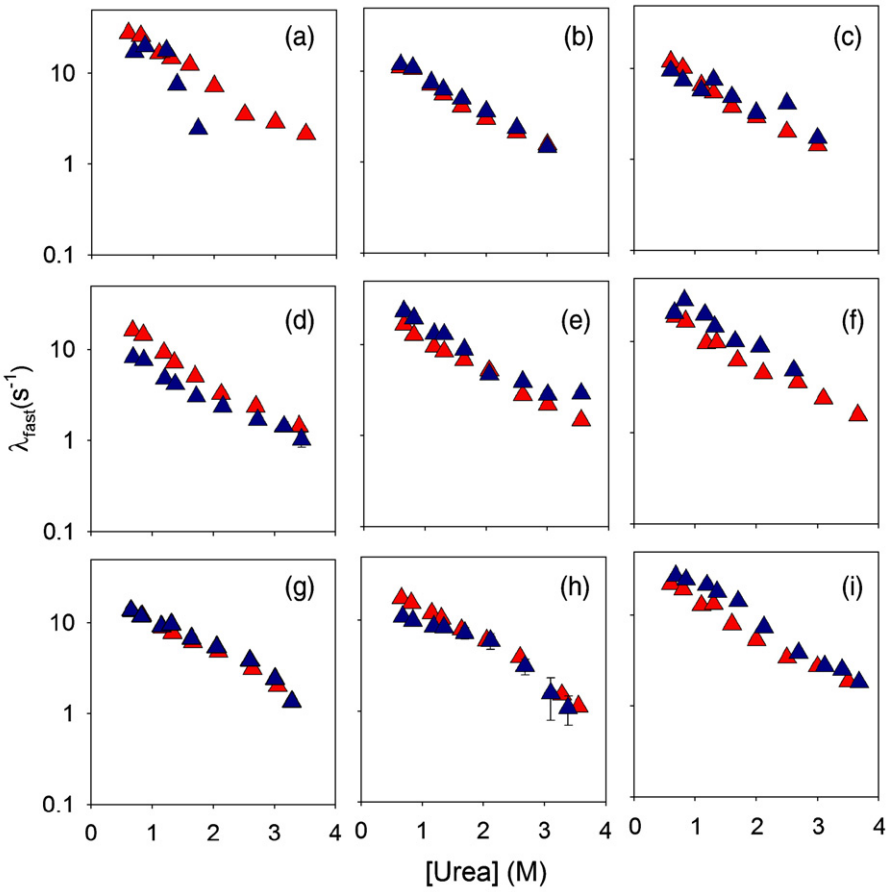

Figure 3. Observed fast refolding rate constants for the different labeled and unlabeled proteins at 25 °C. (a) Cys3 and Cys3-TNB, (b) Cys14 and Cys14-TNB, (c) Cys36 and Cys36-TNB, (d) Cys40 and Cys40-TNB, (e) Cys42 and Cys42-TNB, (f) Cys67 and Cys67-TNB, (g) Cys79 and Cys79-TNB, (h) Cys82 and Cys82-TNB, (i) Cys89 and Cys89-TNB. The red triangles represent the fast refolding rate for the unlabeled protein, and the blue triangles represent the fast refolding rate for the corresponding TNB-labeled proteins. Where shown, error bars represent the standard deviations determined from three independent experiments.

unfolded protein baseline represents the signal of the unfolded protein in that concentration of urea, then the specific component of the burst phase amplitude accounts for 52% (Cys3-TNB), 51% (Cys14-TNB), 40% (Cys36-TNB), 50% (Cys40-TNB), 33% (Cys42-TNB), 41% (Cys67-TNB), and 31% (Cys89-TNB), of the amplitude of refolding from the unfolded protein in refolding condition to completely refolded protein in 0.6 M urea.

As is seen in Figure 4, the specific component of the sub-millisecond folding reaction is seen only at lower concentrations of urea (<2 M) in the form of a deviation from the linearly extrapolated unfolded protein baseline. At higher concentrations of urea (>2 M) in the folding transition region, the $t=0$ points fall on the linearly extrapolated unfolded protein baseline, indicating that only the non-specific component occurs at these concentrations of urea. In fact, this non-specific component is seen for all the labeled proteins at concentrations of urea greater than 2 M, corresponding to marginally stabilizing folding conditions. Thus, the non-specific component arising from the response of the unfolded polypeptide chain to a change in solvent from good to bad is present at higher concentrations of urea for all the proteins, and at all concentrations of urea for Cys25-TNB, Cys62-TNB, Cys79-TNB and Cys82TNB. The slope of the unfolded protein baseline, which is colinear with the linear part of the dependence of the $t=0$ signal on the concentration of urea, is the same (0.065(±0.005) M^{-1}) for all the proteins. Indeed, this is what is expected if the linear dependence on the concentration of urea of the $t=0$ fluorescence intensities extrapolated from the kinetic curves, for folding of the labeled proteins in concentrations of urea greater than 2 M, represents a non-specific effect.

FRET efficiency in the burst phase product varies continuously and differently for different donor-acceptor (D-A) pairs

The data in Figure 4 were used to calculate the FRET efficiency in the product of sub-millisecond folding, for each of the intra-molecular distances being monitored in the individual proteins. The FRET efficiency was calculated using equation (2). The $t=0$ fluorescence signal obtained by extrapolation

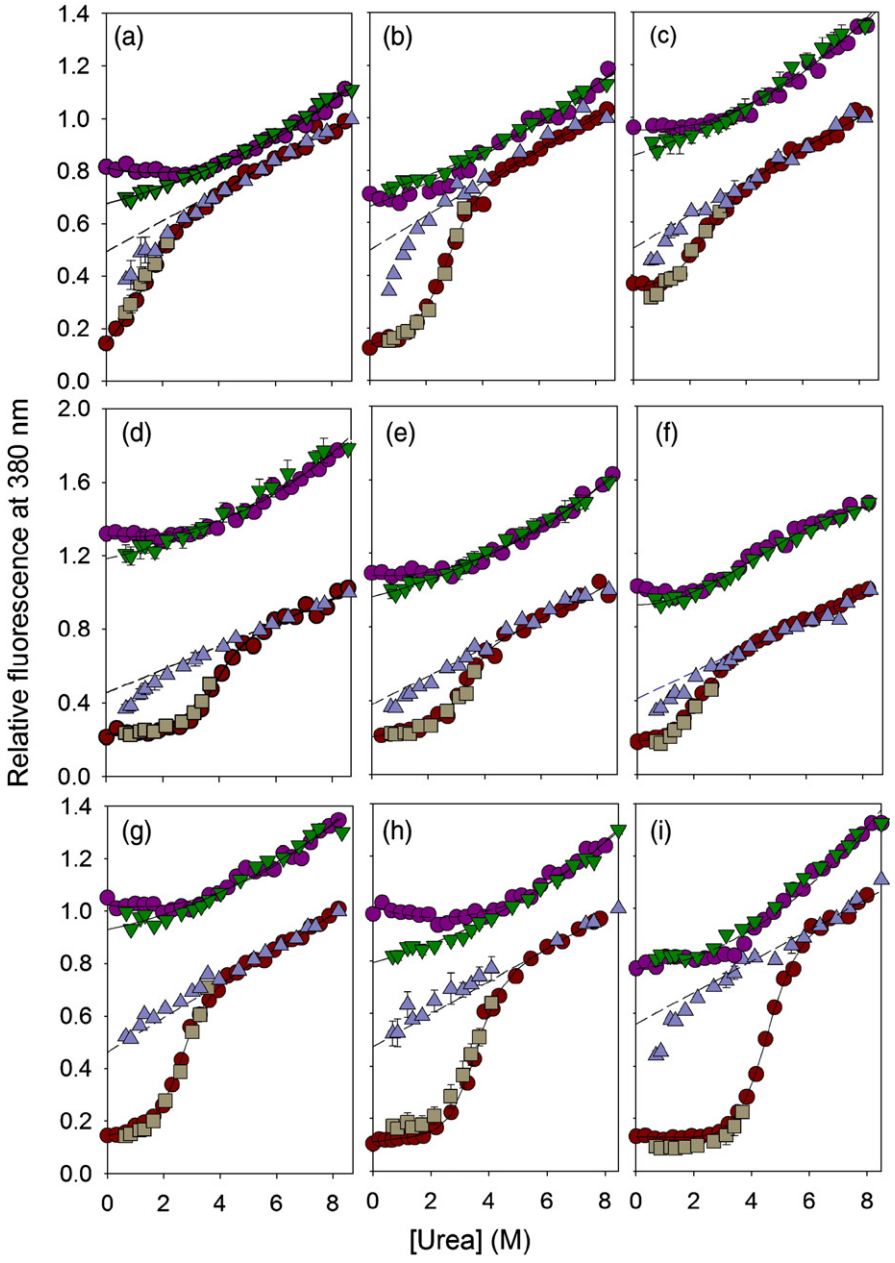

Figure 4. Kinetic *versus* equilibrium amplitudes of the refolding of unlabeled and labeled proteins. (a) Cys3 and Cys3-TNB, (b) Cys14 and Cys14-TNB, (c) Cys36 and Cys36-TNB, (d) Cys40 and Cys40-TNB, (e) Cys42 and Cys42-TNB, (f) Cys67 and Cys67-TNB, (g) Cys79 and Cys79-TNB, (h) Cys82 and Cys82-TNB, (i) Cys89 and Cys89-TNB. In each of the panels, the dark pink circles represent the equilibrium unfolding transition of the unlabeled protein monitored at 380 nm, and the inverted triangles (green) represent the $t=0$ points of the kinetic traces of refolding of the unlabeled protein. The lines through the equilibrium and the kinetic data are second-order polynomial fits. The dark red circles represent the equilibrium unfolding transition of the TNB-labeled protein monitored at 380 nm, and the continuous line through the equilibrium unfolding transition represents a non-linear least-squares fit to a two-state N⇌U model. The values of C_m obtained from the two-state analysis are: 1.3 M, 2.6 M, 2.3 M, 3.8 M, 3.2 M, 1.9 M, 2.6 M, 3.5 M and 4.5 M, for Cys3-TNB, Cys14-TNB, Cys36-TNB, Cys40-TNB, Cys42-TNB, Cys67-TNB, Cys79-TNB, Cys82-TNB, and Cys89-TNB, respectively. The values of C_m obtained for the corresponding unlabeled proteins are 3.8 M, 3 M, 2.8 M, 3.2 M, 3.2 M, 3.7 M, 3.1 M, 2.5 M, 3.6 M and 3.7 M, respectively. Thus, labeling does not alter stability significantly, except in the case of Cys3-TNB. The blue triangles and the squares represent the $t=0$ signal and the $t=\infty$ signal, respectively, obtained from fitting the kinetic traces of refolding of the labeled proteins. The broken line is a linear extrapolation of the unfolded protein baseline. All the data points are normalized to a value of 1 for the signal of the unfolded TNB-labeled protein in 8 M urea.

of the kinetic trace of refolding of the unlabeled protein was taken as F_D (fluorescence intensity at 380 nm of the donor), and the $t=0$ signal from the kinetics trace of refolding of the corresponding TNB-labeled protein was taken as F_{DA} (fluorescence intensity at 380 nm of the donor in presence of the

acceptor). Similarly, for calculation of the FRET efficiency in the species that is populated in different concentrations of urea of the unfolded baseline, the values of the signal obtained immediately after denaturant dilution following stopped-flow mixing (which is essentially the same as the equilibrium value in that particular concentration of urea) for the labeled and the unlabeled proteins were used as F_{DA} and F_D, respectively. It is possible to use the kinetic data obtained for the unlabeled and labeled proteins in this manner to determine the FRET efficiencies, as well as to compare the FRET efficiencies with each other, because each of the individual proteins, labeled as well as unlabeled, for which the FRET data was obtained, display similar folding kinetics and stabilities (Figure 3, and see above). Figure 5(a)–(j) show the dependences of the FRET efficiencies on the concentration of urea in which folding was carried out. For all the proteins, the FRET efficiency appears to have an exponential dependence on the concentration of denaturant. Upon extrapolation of the exponential dependence to zero denaturant, it is seen that the FRET efficiency in the burst phase product is either the same or less than the FRET efficiency in the native protein in the absence of denaturant. In the case of Cys14-TNB (Figure 5(b)), Cys36-TNB (Figure 5(c)) as well as of Cys89-TNB (Figure 5(i)) the extrapolation appears to exceed the value of FRET efficiency calculated for the native state, but it should be remembered that a ~10% error is possible in the calculation of the FRET efficiency in the native state.[56]

Figure 6 compares the dependence on the concentration of urea of the FRET efficiency in the product(s) of sub-millisecond folding, for all the 11 D-A pairs studied so far. The fractional change in efficiency was calculated using equation (6). It is clear from Figure 6(a) that the FRET efficiencies

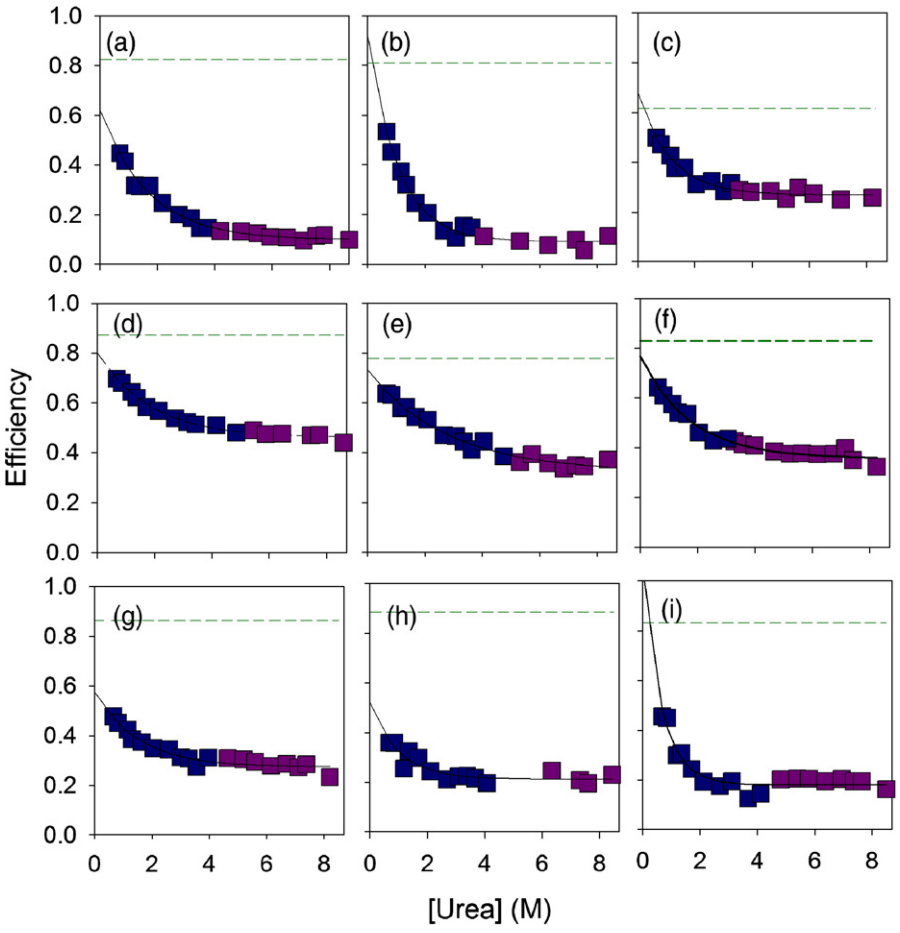

Figure 5. Dependence on urea concentration of the FRET efficiency in the product of sub-millisecond folding. (a) Cys3-TNB, (b) Cys14-TNB, (c) Cys36-TNB, (d) Cys40-TNB, (e) Cys42-TNB, (f) Cys67-TNB, (g) Cys79-TNB, (h) Cys82-TNB and (i) Cys89-TNB. In each panel, the dark pink squares represent the FRET efficiency in the completely unfolded protein forms that are present in different concentrations of urea in the unfolded protein baseline of the equilibrium unfolding transition. The blue squares represent the FRET efficiency in the forms that are populated before the commencement of the fast refolding phase. The continuous line through the efficiency data represents a fit to a single-exponential equation: $E = E(0) + a\, e^{-m[D]}$, where $E(0)$ is the FRET efficiency in 0 M urea, a is the amplitude, and m is the constant that has been used as a measure of the dependence of the FRET efficiency (E) on the concentration of urea ([D]). The broken line represents the FRET efficiency in the native state. The arithmetically propagated standard deviations are ~10% of the mean FRET efficiency.

calculated for the different D-A pairs have different dependences on the concentration of urea. The extents to which the different D-A pairs would have contracted in 0 M urea, as the fits through the data suggests, are also different. The dependence parameter (the value of the constant *m* obtained from the single-exponential fit to the efficiency *versus* concentration of urea data) further highlights the fact that the dependences are uncorrelated (Figure 6(b)): the different D-A pairs do not have the same value of *m*.

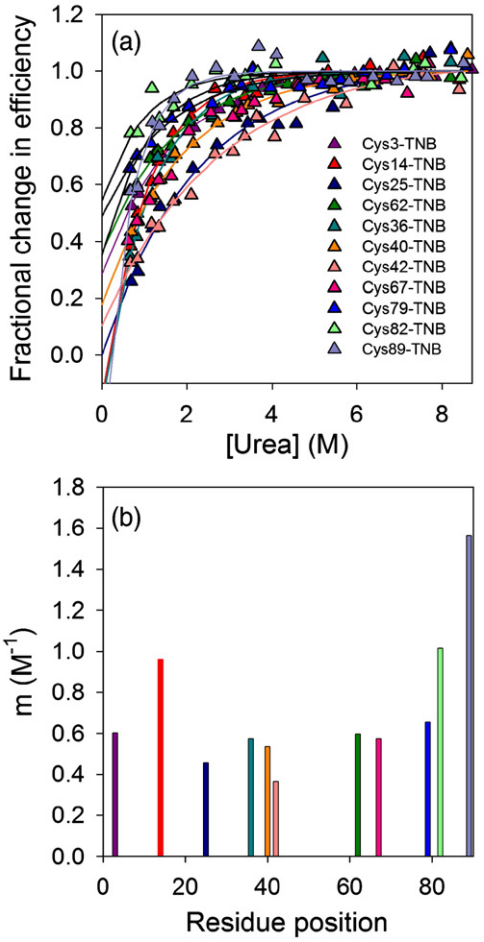

Figure 6. Comparison of the urea-dependence of the FRET efficiencies in the sub-millisecond burst phase product for different D-A pairs. (a) The fractional change in FRET efficiency. The colored triangles represent the fractional change in the FRET efficiency in the initially collapsed form with increasing concentration of urea. Triangles of different colors have been used to show the change in FRET efficiency for different D-A pairs. The fractional change in FRET efficiency was calculated using equation (6). The colored lines represent the fractional change in efficiency calculated using the values from the exponential fits to the efficiency data in Figure 5. Data for Cys25-TNB and Cys62-TNB are taken from Sinha & Udgaonkar.[66] (b) Comparison of dependence obtained from FRET efficiency *versus* [D] plots for different D-A pairs. The parameter *m* (obtained from Figure 5) is plotted against the position of the corresponding acceptor (-Cys-TNB group) in the linear amino acid sequence.

Spectroscopic properties of the burst phase product resemble those of the unfolded protein

An earlier study involving two mutant variants of barstar with solvent-exposed cysteine residues had shown that the values of the quantum yield and the overlap integral (*J*) in the initial burst phase species remain similar to that for the unfolded protein.[66] Figure 7 shows the emission and absorption spectra of the burst phase species populated during the initial few milliseconds of refolding in 0.6 M urea of a mutant protein with a completely buried cysteine residue (Cys40). It is seen in Figure 7(a), that the emission spectrum of tryptophan in the burst phase species is similar to that of the unfolded state. This suggests that the quantum yield of the donor Trp53 is the same in the burst phase product and the unfolded protein.

Figure 7(b) shows the absorption spectrum of the acceptor TNB moiety in the burst phase species. It is seen that this absorption spectrum of the acceptor is blue-shifted when compared to the absorption spectrum in the unfolded protein. Thus, there is some burial of the thiol-TNB group in the burst phase product. This is likely to result in some change in the value of *J*: about a 15% increase is estimated. This is, however, unlikely to cause any significant change in the value of R_0 in the burst phase product, compared to that calculated for the unfolded form, as R_0 has a sixth-root dependence on the value of *J* (equation (4)). Hence, the value of R_0 determined for U has been used for the calculations of the intramolecular distances separating D and A in the burst phase product.

The value of R_0 was calculated separately for each of the D-A pairs studied (Table 1). Since the variation observed in the values of R_0 was small, a mean value of R_0 was used for the calculation of the distances corresponding to the different D-A pairs. The values of R_0 used were 27.0(±0.5) Å for the native state, and 22.5(±0.5) Å for the unfolded state.

D-A distances in the burst phase product contract gradually with decreasing urea concentration

Figure 8 shows the denaturant dependence of the distances that separate donor and acceptor in the product of sub-millisecond folding, for the different mutant proteins. The D-A distances were calculated using equation (3). As discussed above, the value of R_0 for the protein unfolded in 8 M urea was used to calculate the D-A distances in the burst phase species. The data from Figure 5 were used for the values of the efficiency of FRET in the burst phase product. It is evident from Figure 8 that the D-A distances change gradually with the concentration of urea. Like the efficiency *versus* urea concentration data for the different proteins, the dependence of the D-A distances on the concentration of urea are not the same for the different mutant proteins (Figure 9(a) and (b)). In the cases of Cys36-TNB (Figure 8(c)), Cys40-TNB (Figure 8(d)), Cys42-TNB (Figure 8(e)),

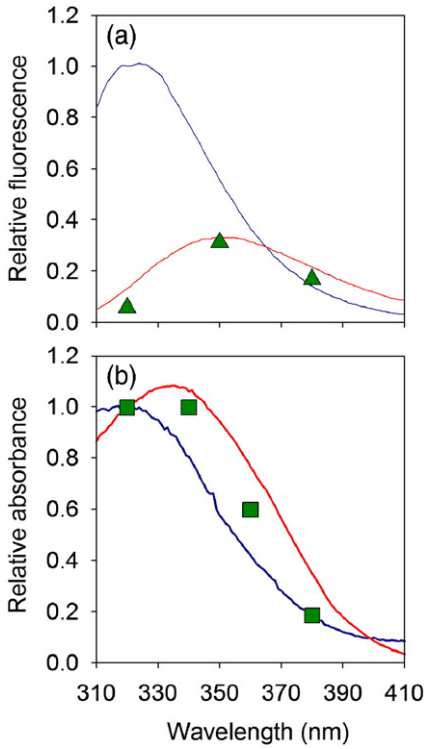

Figure 7. Spectral properties of the burst phase species. The data shown were obtained for Cys40-TNB in which the cysteine residue is >95% buried. (a) The emission spectrum and (b) the absorption spectrum of the burst phase species populated after a few milliseconds of refolding. In (a) the blue line represents the fluorescence emission spectrum of the native state, and the red line represents the spectrum of the unfolded form in 8 M urea. The green upright triangles represent the three-wavelength spectrum of the burst phase species populated during the initial few milliseconds after the initiation of refolding in 0.6 M urea. In (b) the blue line represents the absorbance spectrum of the TNB group in the native protein, and the red line represents the same for the unfolded protein in 8 M urea. The green squares represent a four-wavelength absorbance spectrum of the burst phase species.

Cys67-TNB (Figure 8(f)) and Cys89-TNB (Figure 8(i)), the values determined for the D-A distances in <1 M urea appear to be less than the D-A distance in the native state. It should be noted, however, that the application of FRET methodology to a native protein may have intrinsic problems.[65] In an earlier time-resolved FRET study, an error of about a 10% was found to be associated with the measurement of the distances in the native state.[53] Apart from this, in all these calculations a value of 2/3 has been used for the orientation factor (κ^2) assuming that the dynamic averaging of all the possible orientations of donor and acceptor molecules takes place within the fluorescence lifetime. The range of κ^2 determined for the dansyl group in earlier work[56] was found to be 0.25–2.3 for the native state. This range in the value of κ^2 can give rise to a possible error of ~23% in the calculation of the D-A distances. For TNB, this range is likely to be less, given that it is smaller than dansyl fluorophores.[56] It should be noted that the donor (Trp53) is completely solvent-exposed in the product of sub-millisecond folding,[66] and that, even when attached to a buried cysteine, the TNB label in the product of sub-millisecond folding is as solvent-exposed as it is in the unfolded protein (Figure 7). Hence, the assumption of a value of 2/3 for κ^2 in the product of sub-millisecond folding appears justified.

The root-mean-squared distance ($\langle r^2\rangle^{1/2}$) between the donor and the acceptor groups was determined from the probability distribution function $P(r)$ of a Gaussian-distributed random coil.[67] Briefly, $P(r)$ was used to generate a probability distribution function, $P(E)$ for the FRET efficiency, from which the mean efficiency was determined. $\langle r^2\rangle^{1/2}$ corresponded to the value of the distance for which the mean efficiency so determined was equal to the experimentally determined value of efficiency.[68] It was found (data not shown) that at all concentrations of urea, the values obtained for $\langle r^2\rangle^{1/2}$ exceeded the values of D-A distances calculated directly from the experimentally determined values of the FRET efficiency. But importantly, the dependence of the D-A distance on the concentration of urea remained the same, irrespective of which of the two methods was used to determine distances.

The Gaussian-distributed random coil model has been used widely to describe the dimensions of unfolded proteins,[68–73] and some of these studies report significant deviations of the experimental or simulated data from a Gaussian distribution function.[4,69,70,74,75] When excluded volume effects are explicitly taken into account, a narrower distribution function is found to better describe experimental hydrodynamic data.[75] In fact, it is not known whether a polypeptide chain of finite length behaves like a freely jointed chain under any condition. The Gaussian distribution random coil model appears to be unsuitable for describing the dimensions of an unfolded protein, and its applicability for estimating intra-molecular distances within a collapsed globule, such as the burst phase product of the folding reactions reported here, is even more questionable. It is because of such uncertainties that the D-A

Table 1. Values of R_0 determined for FRET in the native (N) and unfolded (U) forms of different mutant proteins

Protein	R_0 (N) (Å)	R_0 (U) (Å)	D-A distance (N) (Å)
Cys3-TNB	26.6	23.0	20.9
Cys14-TNB	27.3	22.4	21.3
Cys36-TNB	26.4	22.6	25.0
Cys40-TNB	27.0	22.5	19.6
Cys42-TNB	27.0	21.5	22.0
Cys67-TNB	26.9	22.5	20.8
Cys79-TNB	26.7	22.1	20.0
Cys82-TNB	27.9	22.7	19.2
Cys89-TNB	27.8	23.3	20.6

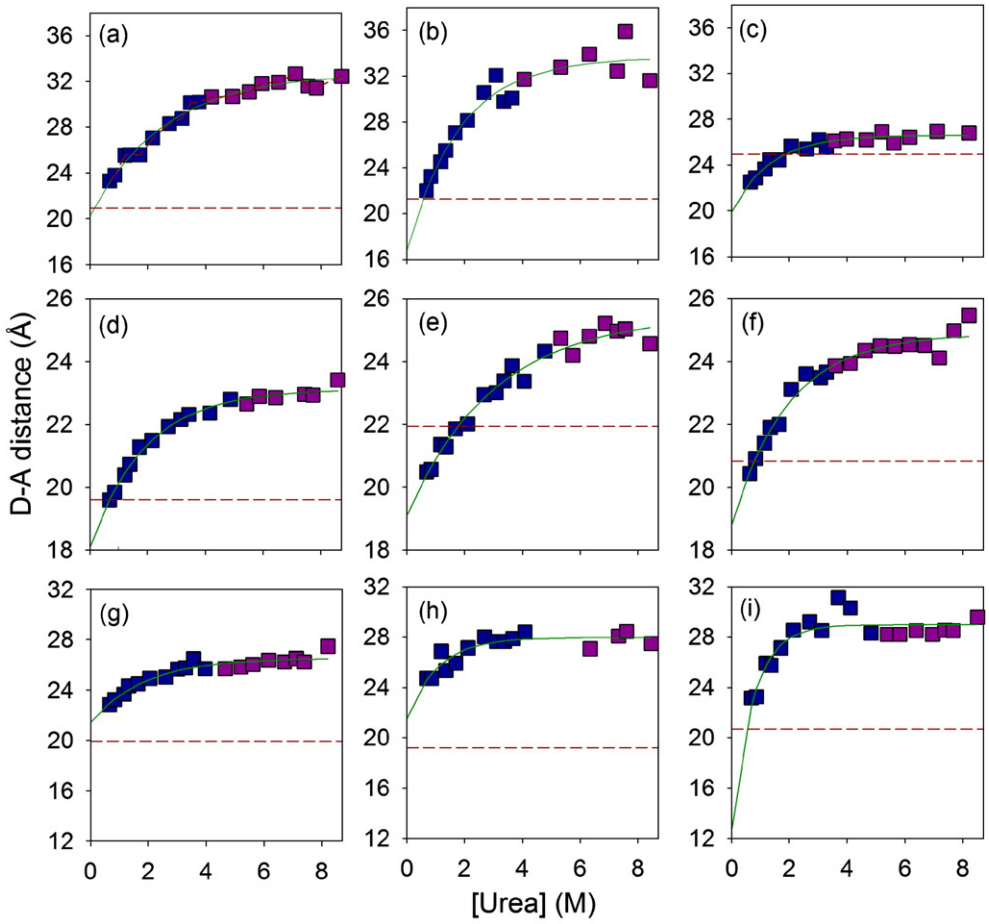

Figure 8. Contraction of intra-molecular distances in the products of sub-millisecond folding. (a) Cys3-TNB, (b) Cys14-TNB, (c) Cys36-TNB, (d) Cys40-TNB, (e) Cys42-TNB, (f) Cys67-TNB, (g) Cys79-TNB, (h) Cys82-TNB, (i) Cys89-TNB. In each panel, the dark pink squares represent the average D-A distances in the unfolded forms in different urea concentrations in the unfolded baseline regions of the equilibrium unfolding transitions. The blue squares represent the average D-A distance in the collapsed forms, which populate the folding pathway before the observed fast refolding phase, in different concentrations of urea. In each panel, the continuous line through the data represents a fit to a single-exponential equation. The broken line represents the D-A distance in the native state (see Table 1).

distances were calculated directly from the experimentally determined values of the FRET efficiency, as described above.

Unstructured peptide fragments do not contract

Figure 10 shows the results of the equilibrium FRET experiments carried out with the peptide fragments 38-55 and 51-69, which represent amino acid residue segments 38–55 and 51–69 in the sequence of Cys40 and Cys67, respectively. The fragments 38-55 and 51-69, after labeling with TNB, contain the Trp53-Cys40TNB and Trp53-Cys67TNB D-A pairs, respectively. The circular dichroism spectra of both of these fragments, in the labeled as well as unlabeled forms, show that they are unstructured in water (data not shown). It is seen from Figure 10 that for both the peptides, the fluorescence intensities of the unlabeled, as well as of the TNB-labeled, peptide increase linearly with the concentration of urea. This indicates that the curvature of the $t=0$ points observed in lower concentrations of urea in the case of the proteins with the Trp53-Cys40TNB (Figure 4(d)) and Trp53-Cys67TNB (Figure 4(f)) D-A pairs, is not because of changing spectral properties of the donor or the acceptor group. It indicates that the non-linear dependence on the concentration of urea of the $t=0$ points extrapolated from the kinetic traces of refolding of Cys40-TNB, Cys67-TNB and other proteins does not arise due to non-specific effects, but must instead represent a specific event. The FRET efficiencies calculated using these data do not show any change with a change in the concentration of urea, which implies that the distance separating Trp53 and Cys40-TNB, as well as Trp53 and Cys67-TNB, does not contract upon dilution of urea, in the context of the peptides. The values of the FRET efficiencies calculated for the distances Trp53-Cys40TNB and Trp53-Cys67TNB in the peptides are about the same as for the corresponding unfolded proteins. In the case of the Trp53-Cys40TNB distance, the values of the FRET efficiency are 0.46 and 0.61 for the unfolded protein and the corresponding peptide fragment

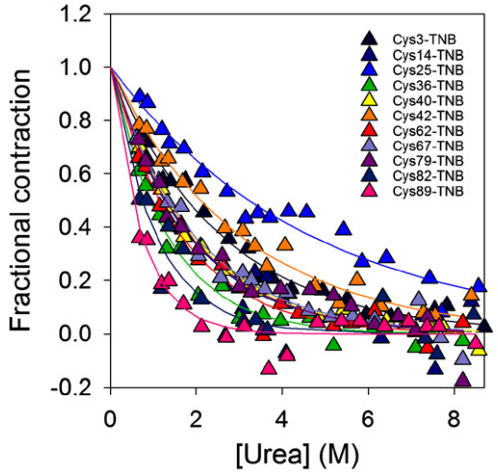

Figure 9. Comparison of the urea dependence of the contraction in the product(s) of sub-millisecond folding for different D-A distances. The dependence on the concentration of urea of the fractional contraction for different D-A pairs is shown as triangles of different colours. The fractional contraction of each distance was calculated using equation (7). The colored lines through the data points represent the fractional contraction calculated using the values obtained from the exponential fit to the experimental data in Figure 8. Data for Cys25-TNB and Cys62-TNB are taken from Sinha & Udgaonkar.[66]

in 8 M urea, respectively. In the case of Trp53-Cys67TNB distance, the values of the FRET efficiency are 0.36 and 0.44 for the unfolded protein and the corresponding peptide fragment in 8 M urea, respectively.

Discussion

Earlier work based on fluorescence and circular dichroism measurements[26,27,60] had suggested that the sub-millisecond folding reaction of barstar involves two components, a non-specific component that results in the formation of a structureless globule, U_C, and a specific component that leads to the formation of a specific intermediate, I_E. It appeared that U_C and I_E, exist in a rapid pre-equilibrium with U during the first few milliseconds of refolding.[27] The relative proportions of the two varies depending on the conditions employed; in marginally stabilizing conditions, U_C is the predominant species present; whereas in the strongly stabilizing conditions, I_E is the major species present. In this study, the non-specific and specific components of the sub-millisecond folding reaction have been identified by the measurement of nine intra-molecular distances using multi-site FRET. Such identification has allowed the cooperativity and specificity of the sub-millisecond folding reaction to be characterized.

Solvent-induced contraction constitutes the non-specific component of sub-millisecond folding

The data in Figure 4 (see Results) show that two of the nine measured intra-molecular distances, Trp53-Cys79TNB and Trp53-Cys82TNB, undergo only non-specific contraction during the first few milliseconds of folding in all concentrations of urea studied. In a previous study, two other intra-

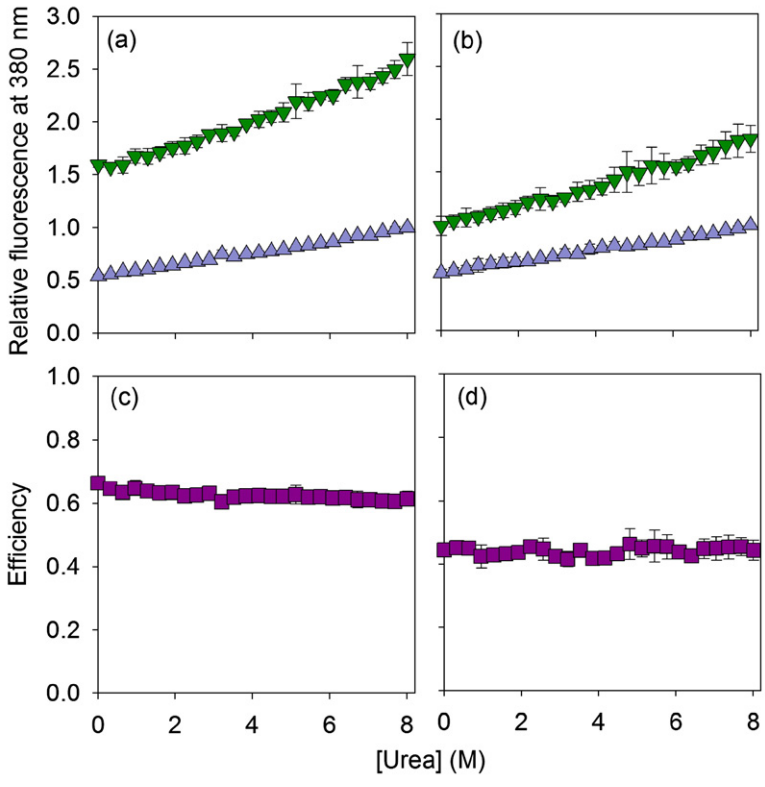

Figure 10. Urea dependence of the fluorescence intensity for two unstructured peptide fragments; (a) fragment 38-55, and (b) fragment 51-69. The green inverted triangles represent the fluorescence signals of the unlabeled peptide, and the blue upright triangles represent the fluorescence intensities of the TNB-labeled peptides, in different concentrations of urea. The FRET efficiency was calculated from the data in (a) and (b) using equation (2). The change in FRET efficiency for (c) fragment 38-55, and (d) fragment 51-69, with concentration of urea, is shown as pink squares. The error bars represent the spread obtained from two independent experiments.

molecular distances, Trp53-Cys25TNB and Trp53-Cys62TNB, were observed to undergo similar non-specific contraction.[66] The extent of contraction is predicted by the response of completely unfolded protein to a reduction in the concentration of urea: the dependence of the burst phase decrease in the fluorescence intensity on urea concentration is identical with, and is colinear with, the urea dependence of the change in fluorescence intensity of the completely unfolded protein. Because the dependence of the signal of the initially collapsed form on the concentration of urea is the same as that of the unfolded protein, it is most likely that this sub-millisecond decrease in the signal results from the response of the unfolded polypeptide chain to the change in solvent (concentration of urea) that is used to initiate refolding. Hence, these four distances undergo only solvent-induced contraction during the first few milliseconds of refolding in strongly stabilizing conditions (concentrations of urea <2 M) as well as in marginally stabilizing conditions (concentrations of urea >2 M). Such solvent-induced contraction of intra-molecular distances constitutes the non-specific component of sub-millisecond folding.

The other seven intra-molecular-distances also show this non-specific component of sub-millisecond refolding, but only in marginally stabilizing conditions (concentrations of urea >2 M). Thus, all intra-molecular distances undergo only solvent-induced contraction during the first few milliseconds of refolding, when the protein is refolded in concentrations of urea greater than 2 M. It has been shown that for folding in such marginally stabilizing conditions, the unfolded protein U collapses initially to form a structure-less globule U_C, which is thought to represent the unfolded form in refolding conditions.[26,27] The results of this study show that all intra-molecular distances have undergone only solvent-induced contraction in U_C, confirming that U_C is the product of non-specific polypeptide chain collapse.

Specific contraction of some chain segments constitutes the specific component of sub-millisecond folding

The Trp53-Cys3TNB, Trp53-Cys14TNB, Trp53-Cys36TNB, Trp53-Cys40TNB, Trp53-Cys42TNB, Trp53-Cys67TNB and Trp53-Cys89TNB distances all contract more than is predicted by the response of unfolded protein to a reduction in the concentration of urea, during the first few milliseconds of refolding in concentrations of urea below 2 M. Previous results had shown that in such more stabilizing conditions, a specific intermediate I_E is populated during the first few milliseconds of refolding.[26,60] I_E is a specific intermediate because its structure is specific to the stability conferred by specific folding conditions,[26,60] and because it possesses secondary structure. The results of this study confirm that I_E is a specific intermediate: out of 11 intra-molecular distances measured, only four have contracted to the extent expected for only solvent-induced contraction. The other seven intra-molecular distances display a specific component of the initial sub-millisecond folding reactions: the extent to which each has contracted in I_E is more than that expected for solvent-induced contraction, and is different for each of the seven distances. This suggests that the extent of specific structure formation in I_E is different in different regions of the molecule. Thus, for refolding in strongly stabilizing conditions, in concentrations of urea below 2 M, the folding transition observed during the first few milliseconds corresponds to $U \rightarrow I_E$. If some of the regions of the protein do not form any specific structure during both the $U \rightarrow U_C$ and $U_C \rightarrow I_E$ transitions in the sub-millisecond time domain, then the contraction of the D-A distances corresponding to those regions would be due solely to the non-specific solvent-dependent effects. Thus, it appears that regions reported on by measurement of the Trp53-Cys25TNB, Trp53-Cys62TNB, Trp53-Cys79TNB and Trp53-Cys82TNB distances are unstructured in U_C and in I_E.

Upon an examination of the positions of the cysteine residues and the corresponding dependence of the burst phase amplitude observed on the concentration of urea, some interesting trends can be seen. The D-A pairs, where the acceptor is placed at a position proximal to the N-terminal side of different helices, viz. Trp53-Cys14TNB, Trp53-Cys36TNB, Trp53-Cys40TNB, and Trp53-Cys67TNB, show the specific component to sub-millisecond folding. The D-A pairs with acceptor located towards the C-terminal side of different helices, viz. Trp53-Cys25TNB, Trp53-Cys62TNB,[66] Trp53-Cys42TNB, and Trp53-Cys79TNB, show only the non-specific component of sub-millisecond folding. This observation suggests that the N-terminal parts of these helices might be more structured than the C-terminal parts, in I_E. This implies that in I_E, the individual helices may be more stable at the N termini, which is surprising, because in an isolated peptide, helix nucleation can occur at any site.[76] On the other hand, little is known about the initiation of helix formation in the context of a protein. Interestingly, the dependence of the burst phase amplitude on the concentration of urea for any of the FRET monitored distances for which the specific component of sub-millisecond folding is seen (Figure 4), is similar to that of the urea-induced unfolding curves of isolated peptide α-helices, which show broad gradual transitions suggestive of non-cooperative unfolding.[77,78]

The specific components of sub-millisecond folding are seen also for two intra-molecular distances where the acceptor is located in a β-strand. In the case of the Trp53-Cys3TNB distance, the acceptor moiety lies close to the N-terminal end of the first β-strand, whereas in the case of Trp53-Cys89TNB distance, the acceptor adduct is on the cysteine residue of the last C-terminal β-strand. Thus, the correlation between the location of the residue and the observation of a specific component of sub-millisecond folding is absent from the case of residues in the β-strands. The acceptor adduct for

the Trp53-Cys82TNB distance lies in a short loop region, and this distance undergoes only non-specific solvent-induced contraction, possibly because the loop region is unstructured in the product of sub-millisecond folding.

Isolated peptides do not collapse

Interestingly, the unstructured peptide fragments corresponding to two of the intra-molecular distances studied, viz. Trp53-Cys40TNB and Trp53-Cys67TNB, do not contract in response to a reduction in the concentration of urea (Figure 10(b)). It is possible that for the initial chain contraction to happen in the protein, some key residues that are distal elsewhere in the sequence, are required to come together during the sub-millisecond folding reaction. It is possible also that for a polymer chain that does not fold into a distinct native structure, there might be a minimum chain length required in order for it to contract. This would be unusual, however, because the persistent length of a polypeptide chain is expected to be the length of two residues.

Sub-millisecond folding leads to formation of non-specific U_C and specific I_E

At present, it is not clear whether the formation of U_C precedes, or occurs in parallel with, the formation of I_E, during initial folding. Since four of the intra-molecular distances undergo only solvent-induced contraction in both U_C and I_E under all folding conditions, and because the remaining seven intra-molecular distances also undergo only solvent-induced contraction under marginally stabilizing folding conditions, it is likely that U_C precedes I_E in the sub-millisecond folding transitions. The $U \rightarrow I_E$ transition might occur in multiple steps *via* one pathway or it might involve multiple pathways. To resolve the temporal sequence of the burst phase events, direct measurement of the kinetics of the sub-millisecond folding reaction is required. Such work is underway.

Nature of initial chain collapse and its product U_C

The initial solvent-induced contraction of the polypeptide chain is a coil to globule transition. For homopolymers, theory predicts that such a coil to globule transition would be steeper for a long polymer chain than for a shorter chain and, consequently, the former would contract relatively more.[79] If the solvent-driven collapse of the unfolded polypeptide chain is like that of a homopolymer, greater contraction should be seen for distances separating the donor and the acceptor groups by a larger number of amino acid residues. Such a correlation is not observed in the present study. For example, the Trp53-Cys3TNB distance contracts less than the Trp53-Cys14TNB and Trp53-Cys25TNB distances, and the Trp53-Cys42TNB distance contracts more than the Trp53-Cys82TNB distance. Thus, the contraction of individual intra-molecular distances in U_C, which correspond to different chain segments of the protein, is different from that expected for a homopolymer chain. The dependence of the extent of contraction on the concentration of urea, is different for all the distances that probe only the non-specific component of the collapse viz. the Trp53-Cys25TNB, Trp53-Cys62TNB, Trp53-Cys79TNB and Trp53-Cys82TNB distances (Figure 9). Differences in the extent of contraction are observed for the other distances for folding in marginally stabilizing conditions (>2 M urea, Figure 9), where U_C is the only product of sub-millisecond folding. It appears, therefore, that the solvent-dependent contraction of a polypeptide chain is non-uniform along its sequence and, hence, is different from that expected for a random polymer chain. The differences imply that the $U \rightarrow U_C$ transition is non-cooperative.[66]

I_E has specific structure and a native-like topology

The value of the FRET efficiency in I_E, can be obtained from the extrapolation to 0 M urea of the single-exponential fit to its dependence on the concentration of urea. For the Trp53-Cys14TNB, Trp53-Cys36TNB, Trp53-Cys40TNB, Trp53-Cys42TNB and Trp53-Cys67TNB distances, the efficiency of FRET in I_E is very similar to that in N. On the other hand, in the case of the Trp53-Cys3TNB, Trp53-Cys79TNB, Trp53-Cys82TNB and Trp53-Cys89TNB distances, the FRET efficiency is intermediate in value between that in N and that in U. The observation that different intra-molecular distances have contracted non-uniformly in I_E, with some having contracted to native-like values, and others distances significantly less, is strongly indicative of I_E possessing specific structure. It appears that I_E has a native-like topology, because many distances in it have contracted to native-like values. Proper topological alignment of different chain segments is expected to be very crucial for the fast protein folding reaction that follows the formation of I_E, as it may decide the probability with which the stabilizing tertiary interactions form.[80]

The specific intermediate I_E undergoes gradual non-cooperative urea-induced unfolding

In this study, the cooperativity of the sub-millisecond refolding reaction has been investigated not by direct measurement of the kinetics, but by measurement of the unfolding by urea, of the product(s) of sub-millisecond refolding. Such an approach, while not ideal, has allowed the use of a multi-site FRET methodology, which permits structural changes in different parts of the protein molecule to be probed. Moreover, measurement of the urea-induced unfolding transitions permits characterization of the cooperativity of the transition from the sub-millisecond folding product to fully unfolded protein, U. As is seen in Figure 5, the FRET

efficiency in the burst phase species changes continuously with a change in the concentration of urea and displays a transition that is not sigmoidal. For all the distances measured, the transition observed for the dependence of the FRET efficiency (Figure 5) and the D-A distance (Figure 8) on the concentration of urea is gradual in nature.

It is perhaps not surprising that that the transition from an early partially folded conformation such as I_E to the unfolded state is gradual in nature. The equilibrium analogs of early kinetic intermediates are the well-known molten globules, and the thermal unfolding of the molten globule forms of several proteins is not two-state but gradual.[81] The NMR-monitored, denaturant-induced unfolding of the molten globule forms of α-lactalbumin and RNase H is non-cooperative.[82,83] The compact denatured state of Engrailed homeodomain, which has been shown by NMR to be a structured intermediate, undergoes progressive unfolding.[84,85] The salt-induced collapse of the high-pH unfolded form of barstar is a non-cooperative higher-order transition.[86,87] In single-molecule FRET studies of several proteins, where the shift in the distribution of the efficiency in the unfolded state with increasing concentration of denaturant was analyzed, the transition from the extended to the compact conformation appeared to be continuous.[68,71,73] It now appears that even a native protein might unfold through a combination of first-order and higher structural transitions.[56] In fact, it has been suggested that the synchrony observed in folding studies that typically employ one or two probes that report on changes only in global properties, is just the average outcome of several asynchronous microscopic folding trajectories.[88] But asynchronous folding or unfolding of different regions of a partially unfolded, kinetic intermediate, at fine structural resolution, has not been observed, until now.

Formation or dissolution of structure in U_C and I_E is not synchronized across different structural regions

Figure 6 shows that the dependence of FRET efficiencies on the concentration of urea for the 11 different D-A pairs are highly uncorrelated; consequently, different distances contract to different extents, relative to the total contraction expected when the molecules undergo the transition from the U to the N state. From Figure 6, it is evident that compaction of different structural segments in the collapsed species is asynchronous. If the observed urea dependence of the FRET efficiency is considered to be a measure of the stability of the structures present in different chain segments, then from Figure 6 it can be inferred that different parts of the unfolded polypeptide chain have not contracted in a synchronous manner in the initially collapsed form. This again means that the initial chain collapse occurs in many steps and is clearly not cooperative. The observation of asynchronous collapse of the polypeptide chain of barstar, across different segments of the chain, is in contrast to the apparently synchronous collapse observed in single-molecule studies of the folding of the small cold-shock protein Csp*Tm*.[73]

The results presented here emphasize the pertinent point that the use of a single probe to measure the cooperativity of the folding reaction can be misleading. Even a sensitive method such as FRET can give misleading results if used to measure only one intra-molecular distance, which might not be representative of structural transitions in the entire molecule. For example, in the case of cytochrome *c*, early work utilizing time-resolved FRET with a single D-A pair had led to the conclusion that the transition between the compact and extended forms of the unfolded polypeptide chain is two-state,[89] but when multiple FRET pairs were used to map the process, intermediate species in the population were seen.[64] In the present study also, if only the Trp53-Cys82TNB distance had been studied, for example, the product of sub-millisecond folding would have appeared to be non-specific in nature. It is only because many different intra-molecular distances, which report on different structural regions, have been monitored by FRET, that it has been possible to obtain a full picture of the lack of cooperativity as well as of synchrony in the sub-millisecond folding reaction.

The transitions leading to the formation of U_C and I_E are continuous transitions

The results from the present multi-site FRET study suggest that the initial sub-millisecond folding reactions of barstar could be processes defined by many small ($\sim k_BT$) barriers. This type of folding has been referred to as downhill folding.[90] This is indicated by the continuous nature and non-coincidence of the observed transitions monitored by measurement of different intramolecular distances, as is seen in Figures 5 and 8. Earlier studies too had indicated that the sub-millisecond folding transition is non-cooperative,[27] and gradual in nature.[66]

As is evident from Figures 5 and 8, there is significant compaction of the chain dimensions in both U_C and I_E. But there is considerable evidence that both U_C and I_E are compacted only loosely, with the core remaining hydrated.[27,59,60,66,87] Hence, their formation would entail little entropic cost, and will be rapid. The large entropic barrier to folding should be present only at later stages of folding when water is extruded from the core, and side-chains are immobilized.[86,91]

The role of polypeptide chain collapse in modulating subsequent structure formation is yet to be understood. Diffusion-driven collisions of segments of the polypeptide chain may not be avoidable in the more dense globule produced by the collapse, and the consequent internal friction[92] may slow folding. On the other hand, if the diffusive motions of the chain during folding are slaved to solvent fluctuations,[93] it is not clear what the effect of chain collapse might be on folding

kinetics. It is possible that the reduction of conformational space in the collapsed globule, which is available for the diffusing chain to sample, might indeed speed folding in the globule. All enthalpically stabilized conformations will be sampled rapidly by the diffusing chain within the globule, and these would include many with non-native interactions. Collapse is expected to facilitate formation of sub-structures such as secondary structural units, which are themselves compact. If hydrogen bonded water is extruded from the vicinity of the chain, then secondary structure formation will be further facilitated,[94] and the entropic cost of structure formation may be paid for by the entropic gain resulting from the release of water molecules.

Can specific structure form *via* a continuous transition?

In the present study, evidence for the specificity of the structure in I_E comes from the observation that different distances have contracted to different extents (Figures 8 and 9). At the same time, measurement of the expansion of the 11 intra-molecular distances upon an increase in the concentration of urea indicates that both the $U \rightarrow U_C$ and $U \rightarrow U_C \rightarrow I_E$ transitions are non-cooperative and gradual in nature. If I_E is indeed formed by a gradual diffusive process, then its structure is expected to be different under different folding conditions, because all intermediate structures would be separated by very small energy barriers. In this study, the observation that different intra-molecular distances in I_E have contracted differently in response to a reduction in the concentration of urea, indicates that the structure present in I_E is different in the different concentrations of urea in which folding was carried out. In other words, the structure of I_E is specific to the specific conditions utilized for folding. Similar observations were made when the refolding was carried out in the presence of different salts and osmolytes;[27,60] different structures are observed for I_E under conditions that confer different stability. It appears that due to near-degenerate energies, these different forms can interconvert when the solvent conditions are changed. It would not be too unusual for the specific structure in I_E to form *via* an asynchronous and continuous diffusive process. One example of specific structure forming *via* a conformational diffusion search appears to be the formation of helices from coils: both experiments and simulations have suggested that the coil to helix transition occurs by a diffusive search within the coil.[54,95] The folding of several small proteins also appears to be downhill in nature.[96,97]

Heterogeneity in I_E

It is seen that in I_E, the efficiency of FRET is different for different intra-molecular distances (Figure 6). From steady state FRET measurements, it is not possible to determine whether all of these intra-molecular distances have contracted in all the molecules. For the distances where it is the same as in N, it is obvious that all the molecules in the population must have contracted to native-like dimensions for these distances. In contrast, for the intra-molecular distances characterized by intermediate values of FRET efficiency, such as for Trp53-Cys82TNB, where it is ~50% of the native state value, there are two explanations possible. One possibility is that all the molecules in the population have contracted to I_E such that the FRET efficiency in I_E is half of that in N. This is consistent with I_E possessing specific structure. The other possibility is that only half of the molecules in the population have collapsed such that the FRET efficiency has become the same as that in N, and the remaining half remain in the extended conformations of U. The second possibility would arise only if a barrier is present that prevents half of the molecules from contracting. It implies heterogeneity in I_E, which would consist of compact and extended conformations. This also implies that multiple pathways are available for folding, because the molecules that do not collapse initially also finally transform in the native structure. It should be noted that previous work has indicated that barstar does indeed utilize multiple pathways to fold, which arise, in part, from different unfolded forms.[58,98] Nevertheless, all unfolded molecules have been shown to undergo the sub-millisecond folding reactions,[99] including those that refold slowly.[59]

Do multiple pathways lead to the formation of U_C and I_E?

Upon a rapid change to refolding conditions, every individual conformation of an unfolded protein is expected to follow its own collapse trajectory. If collapse is diffusive, and if different polypeptide chains collapse progressively, do all molecules follow the same pathway of collapse? In other words, is there one specific pathway representing the $U \rightarrow U_C \rightarrow I_E$ transition? Or can different large sets of molecular trajectories each be averaged into a different pathway reporting on a different progression of collapse events? In simulations, preferred folding/unfolding pathways at the macroscopic level are seen as the averaged outcome of multiple folding/unfolding trajectories at the microscopic level.[100,101]

One explanation for the observation (Figure 6) that different segments of the polypeptide chain collapse differently upon a reduction in the concentration of urea is that the collapse occurs in multiple steps. A more likely explanation, give the rather wide dispersion in the dependence on the concentration of urea for the contraction of different distances, is that different pathways are involved in the sub-millisecond folding reactions. This is likely, given that the starting state, the unfolded state, is itself heterogeneous,[4,56] as are the products

of sub-millisecond folding, U_C and I_E.[27,60] There is also evidence from ANS binding experiments for multiple folding pathways,[59] as well as evidence of heterogeneity in the later steps of folding.[63,102] Heterogeneity in a folding intermediate ensemble is likely to be a consequence of multiple routes and pathways.[103] For example, in the case of apomyoglobin and bacteriophage T4 lysozyme, the observed differences in the burst phase protection of different amide protons in different structural regions of the protein were interpreted to be an outcome of heterogeneity in both the structures and the folding pathways.[104,105] It is expected that direct measurements of the sub-millisecond folding kinetics, now in progress, will provide definitive evidence that initial folding occurs along multiple pathways.

Materials and Methods

Protein expression, purification and labeling

Wild type barstar has three tryptophan residues, at positions 38, 44, and 53, and there are two cysteine residues, at positions 40 and 82. The mutant proteins Cys3, Cys14, Cys36, Cys40, Cys42, Cys67, Cys79, Cys82 and Cys89 each contain only a single cysteine residue at the locations indicated in their names, as well as only a single tryptophan residue, Trp53. All the mutant proteins described in this work were generated by site-directed mutagenesis, and the proteins were purified as described.[62,66] All proteins were judged by SDS-PAGE to be >98% pure. The mass of each of the proteins was determined by mass spectrometry using a Micromass Q-TOF Ultima instrument, and was found to be consistent with the mass expected for the mutant protein with the N-terminal methionine residue remaining uncleaved.

Labeling the proteins with TNB was achieved by incubating the urea-unfolded protein with a 20-fold molar excess of 5, 5′-dithiobis (2-nitrobenzoic acid) (DTNB) at pH 8.5. After completion of the reaction, the labeled protein was separated from free dye and urea by the use of a PD-10 column (Pharmacia). All the proteins were found by mass spectrometry to be >95% labeled, with an expected 196 Da increase in the mass due to the TNB adduct.

Buffers, solutions and experimental conditions

All the reagents used were of the highest purity grade obtained from Sigma. Buffers used in all the experiments contained 20 mM Tris and 0.25 mM EDTA. All the experiments were performed at 25 °C. The concentration of the stock urea solution was determined by refractive index measurements on an Abbe refractometer. In all the equilibrium experiments, the protein concentration used was 3–4 μM in the case of the unlabeled proteins, and 8–10 μM in the case of the TNB-labeled proteins.[66] In all the kinetic experiments, the concentration of protein was 5-10 μM in the case of the unlabeled proteins, and 15-20 μM in the case of the TNB-labeled proteins.[66] It was necessary to use a higher concentration for the labeled protein because of the quenching of the fluorescence signal by the TNB label. The protein concentration was determined by measuring the absorbance at 280 nm, using $\varepsilon_{280} = 10000\ M^{-1}cm^{-1}$. Since the TNB group contributes to the absorbance measured at 280 nm, a correction for its contribution was done for the labeled proteins as described earlier.[66]

Acquisition of emission and absorption spectra

The emission spectra of the native and unfolded states were acquired on a Fluoromax-3 fluorimeter. The three-wavelength emission spectrum of the burst phase species was constructed by using the $t = 0$ points obtained from dead-time extrapolation of the kinetic traces of refolding in 0.6 M urea, monitored at the three different wavelengths, using a Biologic SFM-400 stopped-flow instrument.

The absorption spectra for the native and the unfolded proteins were acquired on a Cary 100 double-beam spectrophotometer with a bandwidth of 1 nm and a scan rate of 1 nm/s, using a 1 cm path-length cuvette. The absorption spectrum of the burst phase product was constructed by monitoring the absorbance changes at four different wavelengths after the initiation of refolding in 0.6 M urea on an SFM-4 module using a FC-20 cuvette. For data points at 340 nm and 360 nm, the $t = 0$ points obtained after dead-time extrapolation of the refolding traces were used, whereas at 320 nm and 380 nm, time-averaged absorbance values were used as no process could be observed.

Equilibrium unfolding experiments

All the equilibrium unfolding experiments were carried out on a Fluoromax-3 (Jobin Yvon) spectrofluorimeter, or a Biologic SFM-4 stopped-flow instrument. The protein was incubated in different concentrations of urea for ~3 h and the equilibrium unfolding transition was monitored by exciting tryptophan fluorescence at 295 nm with the slit-width set at 0.4 nm, and by measuring the emission at 380 nm with the slit-width set at 10 nm.

The equilibrium unfolding data were analyzed using a two-state $N \rightleftharpoons U$ model to obtain the values of the free energy of unfolding in water and of C_m, the mid-point of the unfolding transition.[106]

Kinetic refolding experiments

All the kinetic experiments were done on a Biologic SFM-4 stopped-flow instrument with a FC-15 cuvette. The dead-time of measurement was 6.2 ms. Excitation was carried out at 295 nm, and emission was acquired at 380 nm using a 10 nm band-pass filter as described.[66]

To normalize the signals of the labeled and unlabeled proteins, the fluorescence intensities of known concentrations of labeled and unlabeled proteins were measured in 8 M urea under identical experimental conditions.

The average of several (8–10) kinetic traces was fit to a sum of three exponentials:

$$F = F_0 + F_1\, e^{-\lambda_1 t} + F_2\, e^{-\lambda_2 t} + F_3\, e^{-\lambda_3 t} \qquad (1)$$

where F is the fluorescence intensity, which changes as a function of time t, λ_1, λ_2, and λ_3 are the three observed rate constants with amplitudes F_1, F_2 and F_3, respectively.

FRET experiments with peptide fragments

Custom-synthesized peptide fragments of barstar (>95% purity), corresponding to sequence segments 38–55 and 51–69 were obtained from GenScript Corporation. The labeling of the peptides with the TNB adduct was achieved by incubating the peptides dissolved in native buffer (20 mM Tris, 250 μM EDTA, pH 8.0) with the DTNB solution at pH 8.0. The labeled peptide was separated from the free dye using a Mono-S Reverse Phase HPLC column over an 80% (v/v) acetonitrile gradient or by using a GE Hi-Trap desalting (G-25) column, on an AKTA HPLC system. The extent of labeling was checked using mass spectrometry and was found to be >95%. Equilibrium experiments with both the labeled and the unlabeled peptides were performed under identical fluorimeter settings by incubating the peptides in different concentrations of urea for ~1 h. The peptide concentrations used were ~5 μM for the unlabeled peptides, and 15–20 μM for the TNB-labeled peptides. The concentration was determined using the value of the extinction coefficient calculated from the number of tryptophan and tyrosine residues. These values are 7090 $M^{-1}cm^{-1}$ and 5810 $M^{-1}cm^{-1}$ for the fragments 38-55 and 51-69, respectively. The samples were excited at 295 nm, and the emission was collected at 380 nm with slit-widths of 0.4 nm and 10 nm for excitation and emission, respectively. To normalize the fluorescence intensity of the labeled and the unlabeled proteins, the absorbance value measured at 280 nm for the TNB-labeled peptide was corrected for the contribution of the TNB group. The contribution of the TNB group to the absorbance at 280 nm was estimated from its contribution to A_{280} in the corresponding labeled protein.[66]

Data analysis

Analysis of FRET data

The FRET efficiency was calculated using the equation:

$$E = 1 - (F_{DA}/F_D) \quad (2)$$

where F_D is the fluorescence intensity of the donor (measured in the unlabeled protein) and F_{DA} is fluorescence intensity of the donor in the presence of the acceptor (measured in the corresponding labeled protein).

Forster's relationship between the distance R and the energy transfer efficiency E is given by:

$$E = R_0^6/(R_0^6+R^6) \quad (3)$$

R_0 is the Forster's distance characteristic of the D-A pair under study.

The value of R_0 was calculated using following relationship:

$$R_0 = 0.211(Q_D J\kappa^2 n^{-4})^{1/6} \quad (4)$$

Q_D is the quantum yield of the donor, J is the overlap integral that signifies the spectral overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor, κ^2 is the orientation factor, and n is the refractive index of the medium used.

The value of J was determined using:

$$J = \int F(\lambda)\varepsilon(\lambda)\lambda^4 d\lambda / \int F(\lambda)d\lambda \quad (5)$$

The values of Q_D for the native and the unfolded protein are known to be 0.27 and 0.11, respectively.[56] In all the calculations, a value of 2/3 for κ^2 was used,[56,66] which assumes free rotation of the donor and acceptor molecules in all the forms. The value of the refractive index was determined separately for the native and the unfolding buffers.

Calculation of the fractional change in efficiency

The fractional change in FRET efficiency was calculated as:

$$F_E = (E_N - E_C)/(E_N - E_U) \quad (6)$$

where F_E is the fractional change in FRET efficiency, E_N is the efficiency in the native state, E_C is the efficiency in the burst phase species, and E_U is the efficiency in the unfolded state.

The fractional change in the D-A distance was calculated using the following equation:

$$F_C = (D_U - D_C)/(D_U - D_0) \quad (7)$$

where F_C is the fractional contraction, D_U is the D-A distance in the unfolded state (in 8 M urea), D_C is the D-A distance in the collapsed form, and D_0 is the value of the D-A distance in the collapsed form in 0 M urea obtained by extrapolation.

Acknowledgements

We thank G. Krishnamoorthy, M. K. Mathew and the members of our laboratory for discussions. We thank G. Haran for his program for calculating the mean efficiency for a Gaussian chain distribution of intra-molecular distances. K.K.S. is the recipient of a Senior Research Fellowship from the CSIR, India. This work was funded by the Tata Institute of Fundamental Research, and by the Department of Biotechnology, Government of India.

References

1. Tanford, C., Kawahara, K. & Lapanje, S. (1966). Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior. *J. Biol. Chem.* **241**, 1921–1923.
2. Goldenberg, D. P. (2003). Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **326**, 1615–1633.
3. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N. *et al.* (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12491–12496.
4. Saxena, A. M., Udgaonkar, J. B. & Krishnamoorthy, G. (2006). Characterization of intramolecular distances and site-specific dynamics in chemically unfolded barstar: evidence for denaturant-dependent non-random structure. *J. Mol. Biol.* **359**, 174–189.
5. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12497–12502.
6. Mayor, U., Grossmann, J. G., Foster, N. W., Freund, S. M. V. & Fersht, A. R. (2003). The denatured state of

Engrailed homeodomain under denaturing and native conditions. *J. Mol. Biol.* **333**, 977–991.
7. Reed, M. A. C., Jelinska1, C., Syson, K., Cliff, M. J., Splevins, A., Alizadeh, T. *et al.* (2006). The denatured state under native conditions: a non-native-like collapsed state of N-PGK. *J. Mol. Biol.* **357**, 365–372.
8. Anil, B., Li, Y., Cho, J.-H. & Raleigh, D. P. (2006). The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemistry*, **45**, 10110–10116.
9. Uversky, V. N., Gillespie, J. R., Millett, I. S., Khodyakova, A. V., Vasiliev, A. M., Chernovskaya, T. V. *et al.* (1999). Natively unfolded human prothymosin R adopts partially folded collapsed conformation at acidic pH. *Biochemistry*, **38**, 15009–15016.
10. Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2–12.
11. Morar, A. S., Olteanu, A., Young, G. B. & Pielak, G. J. (2001). Solvent induced collapse of α-synuclein and acid-denatured cytochrome c. *Protein Sci.* **10**, 2195–2199.
12. Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825.
13. Ladurner, A. G. & Fersht, A. R. (1999). Upper limit of the time scale for diffusion and chain collapse in chymotrypsin inhibitor 2. *Nature Struct. Biol.* **6**, 28–31.
14. Hagen, S. J. & Eaton, W. A. (2000). Two-state expansion and collapse of a polypeptide. *J. Mol. Biol.* **297**, 781–789.
15. Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. (1997). Ultra fast signals in protein folding and the polypeptide contracted state. *Proc. Natl Acad. Sci. USA*, **94**, 8545–8550.
16. Qi, P. X., Sosnick, T. R. & Englander, S. W. (1998). The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nature Struct. Biol.* **5**, 882–884.
17. Bhuyan, A. K. & Udgaonkar, J. B. (1999). Relevance of burst phase changes in optical signals of polypeptides during protein folding. In *Perspectives in Structural Biology* (Vijayan, M., Yathindra, N. & Kolaskar, A. S., eds), pp. 293–303, Universities Press, Hyderabad.
18. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.* **324**, 359–371.
19. Spector, S. & Raleigh, D. P. (1999). Submillisecond folding of the peripheral subunit-binding domain. *J. Mol. Biol.* **293**, 763–768.
20. Yang, W. Y. & Gruebele, M. (2003). Folding at the speed limit. *Nature*, **423**, 193–197.
21. Elove, G. A., Chaffotte, A. F., Roder, H. & Goldberg, M. E. (1992). Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. *Biochemistry*, **31**, 6876–6883.
22. Yamasaki, K., Ogasahara, K., Yutani, K., Oobatake, M. & Kanaya, S. (1995). Folding pathway of *Escherichia coli* ribonuclease HI: a circular dichroism, fluorescence, and NMR study. *Biochemistry*, **34**, 16552–16562.
23. Karplus, M. & Weaver, D. L. (1979). Diffusion collision model for protein folding. *Biopolymers*, **18**, 1421–1437.
24. Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–489.
25. Ptitsyn, O. B. (1995). Molten globule and protein folding. *Advan. Protein Chem.* **47**, 83–229.
26. Agashe, V. R., Shastry, M. C. R. & Udgaonkar, J. B. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature*, **377**, 754–757.
27. Pradeep, L. & Udgaonkar, J. B. (2004). Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Bio. Chem.* **279**, 40303–40313.
28. Magg, C. & Schmid, F. X. (2004). Rapid collapse precedes the fast two-state folding of the cold shock protein. *J. Mol. Biol.* **335**, 1309–1323.
29. Dumont, C., Matsumura, Y., Kim, S. J., Li, J., Kondrashkina, E., Kihara, H. & Gruebele, M. (2006). Solvent tuning the collapse and helix time scale of $\lambda^*_{6\text{-}85}$. *Protein Sci.* **15**, 2596–2604.
30. Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S. & Kiefhaber, T. (1998). Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J. Mol. Biol.* **288**, 489–499.
31. Uversky, V. N. & Ptitsyn, O. B. (1996). Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* **255**, 215–228.
32. Navon, A., Ittah, V., Landsman, P., Scheraga, H. A. & Haas, E. (2001). Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-Terminal portions of the reduced protein. *Biochemistry*, **40**, 105–118.
33. Sadqi, M., Lapidus, M. J. & Munoz, V. (2003). How fast is protein hydrophobic collapse? *Proc. Natl Acad. Sci. USA*, **100**, 12117–12122.
34. Arai, M., Kondrashkina, E., Kayatekin, C., Matthews, C. R., Iwakura, M. & Bilsel, O. (2007). Microsecond hydrophobic collapse in the folding of *Escherichia coli* dihydrofolate reductase, an α/β-type protein. *J. Mol. Biol.* **368**, 219–229.
35. Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1995). Is burst hydrophobic collapse necessary for protein folding? *Biochemistry*, **34**, 3066–3076.
36. Zagrovic, B., Snow, C. D., Shirts, M. R. & Pande, V. S. (2002). Simulation of folding of a small alfa-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **323**, 927–937.
37. Welker, E., Maki, K., Shastry, M. C. R., Juminaga, D., Bhat, R., Scheraga, H. A. & Roder, H. (2004). Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A. *Proc. Natl Acad. Sci. USA*, **101**, 17681–17686.
38. Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. & Fujisawa, T. (2002). Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc. Natl Acad. Sci. USA*, **99**, 1329–1334.
39. Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T. *et al.* (2005). Specific collapse followed by slow hydrogen bond formation of β-sheet in the folding of single chain monellin. *Proc. Natl Acad. Sci. USA*, **102**, 2748–2753.
40. Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I. & Fujisawa, T. (2004). Collapse and search dynamics of apomyoglobin folding revealed by sub-millisecond observations of α-helical content and compactness. *Proc. Natl Acad. Sci. USA*, **101**, 1171–1176.

41. Ballew, R. L., Sabelko, J. & Gruebele, M. (1996). Direct observation of fast protein folding. *Proc. Natl Acad. Sci. USA*, **93**, 5759–5764.
42. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*, **262**, 892–896.
43. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15–28.
44. Chan, H. S., Bromberg, S. & Dill, K. A. (1995). Models of cooperativity in protein folding. *Phil. Trans. Roy. soc. ser. B*, **348**, 61–70.
45. Parker, M. J. & Marqusee, S. (1999). The cooperativity of burst phase reactions explored. *J. Mol. Biol.* **293**, 1195–1210.
46. Georgescu, R. E., Lee, J. H., Goldberg, M. E., Tasayko, M. L. & Chaffotte, A. F. (1998). Proline isomerization-independent accumulation of early intermediate and heterogeneity of the folding pathways of a mixed alpha/beta protein, *Escherichia coli* thioredoxin. *Biochemistry*, **37**, 10286–10297.
47. Takei, J., Chu, R.-A. & Bai, Y. (2000). Absence of stable intermediates on the folding pathway of barnase. *Proc. Natl Acad. Sci. USA*, **97**, 10796–11801.
48. Shastry, M. C. R. & Roder, H. (1998). Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nature Struct. Biol.* **5**, 385–392.
49. Hagen, S. J. & Eaton, W. A. (2000). Two-state expansion and collapse of a polypeptide. *J. Mol. Biol.* **297**, 781–789.
50. Qiu, L., Zachariah, C. & Hagen, S. J. (2003). Fast chain contraction during protein folding: foldability and collapse dynamics. *Phys. Rev. Letters*, **90**, 168103-1–168103-4.
51. Zhou, H.-X. & Zwanzig, R. (1991). A rate process with an entropy barrier. *J. Chem. Phys.* **94**, 6147–6152.
52. Hagen, S. J. (2003). Exponential decay kinetics in downhill protein folding. *Protein Struct. Funct. Genet.* **50**, 1–4.
53. Sabelko, J., Ervin, J. & Gruebele, M. (1999). Observation of strange kinetics in protein folding. *Proc. Natl Acad. Sci. USA*, **96**, 6031–6036.
54. Huang, C.-Y., Getahun, Z., Zhu, Y., Klemke, J. W., DeGrado, F. & Gai, F. (2002). Helix formation via conformational diffusion search. *Proc. Natl Acad. Sci. USA*, **99**, 2788–2793.
55. Leeson, D. T., Gai, F., Rodriguez, H. M., Gregoret, L. M. & Dyer, R. B. (2000). Protein folding and unfolding on a complex energy landscape. *Proc. Natl Acad. Sci. USA*, **97**, 2527–2532.
56. Lakshmikant, G. S., Sridevi, K., Krishnamoorhty, G. & Udgaonkar, J. B. (2001). Structure is lost incrementally during the unfolding of barstar. *Nature Struct. Biol.* **8**, 799–804.
57. Hagen, S. J., Hofrichter, J. & Eaton, W. A. (1997). Rate of intrachain diffusion of unfolded cytochrome c. *J. Phys. Chem. B*, **101**, 2352–2365.
58. Schreiber, G. & Fersht, A. R. (1993). The refolding of cis- and trans-peptidylprolyl isomers of barstar. *Biochemistry*, **32**, 11195–11203.
59. Shastry, M. C. R. & Udgaonkar, J. B. (1995). The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* **247**, 1013–1027.
60. Pradeep, L. & Udgaonkar, J. B. (2002). Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar. *J. Mol. Biol.* **324**, 331–347.
61. Lillo, M. P., Beechem, J. M., Szpikowska, B. K., Sherman, M. A. & Mas, M. T. (1997). Design and characterization of a multisite fluorescence energy-transfer system for protein folding studies: a steady-state and time-resolved study of yeast phosphoglycerate kinase. *Biochemistry*, **36**, 11261–11272.
62. Sridevi, K. & Udgaonkar, J. B. (2003). Surface expansion is independent of and occurs faster than core salvation during the unfolding of barstar. *Biochemistry*, **42**, 1551–1563.
63. Sridevi, K., Lakshmikanth, G. S., Krishnamoorthy, G. & Udgaonkar, J. B. (2004). Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble. *J. Mol. Biol.* **337**, 699–711.
64. Pletneva, E. V., Gray, H. B. & Winkler, J. R. (2005). Snapshots of cytochrome c folding. *Proc. Natl Acad. Sci. USA*, **102**, 18397–18402.
65. Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy.* Kluwer Academic, Plenum Publishers, New York.
66. Sinha, K. K. & Udgaonkar, J. B. (2005). Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: evidence for a continuous transition. *J. Mol. Biol.* **353**, 704–718.
67. Cantor, C. R. & Schimmel, P. R. (1980). Configurational statistics of polymer chains. In *Biophysical Chemistry*, chapt. 18, W. H. Freeman & Co., New York.
68. Sherman, E. & Haran, G. (2006). Coil-globule transition in the denatured state of a small protein. *Proc. Natl Acad. Sci. USA*, **103**, 11539–11543.
69. Wang, X., Bodunov, E. N. & Nau, W. M. (2003). Fluorescence quenching kinetics in short polymer chains: dependence on chain length. *Optics and Spectroscopy*, **95**, 603–613.
70. Laurence, T. A., Kong, X., Jager, M. & Weiss, S. (2005). Probing structural heterogenieties and fluctuations of nucleic acids and denatured proteins. *Proc. Natl Acad. Sci. USA*, **102**, 17348–17353.
71. Kuzmenkina, E. V., Heyes, C. D. & Nienhaus, G. U. (2006). Single-molecule FRET study of denaturant induced unfolding of RNase H. *J. Mol. Biol.* **357**, 313–324.
72. Moglich, A., Joder, K. & Kiefhaber, T. (2006). End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl Acad. Sci. USA*, **103**, 12394–12399.
73. Hoffmann, A., Kane, A., Nettels, D., Hertzog, D. E., Baumgartel, P., Lengefeld, J. *et al.* (2007). Mapping protein collapse with single-molecule fluorescence and kinetic synchroton radiation circular dichroism spectroscopy. *Proc. Natl Acad. Sci. USA*, **104**, 105–110.
74. Goldenberg, D. P. (2003). Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **326**, 1615–1633.
75. Zhou, H.-X. (2002). Dimensions of denatured protein chains from hydrodynamic data. *J. Phys. Chem. B*, **106**, 5769–5775.
76. Finkelstein, A. V. & Ptitsyn, O. B. (2002). In *Protein Physics,* chapt. 17 and 18, pp. 207–237, Academic Press, London.
77. Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M. & Bolen, D. W. (1991). Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. *Proc. Natl Acad. Sci. USA*, **88**, 2854–2858.

78. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. (1996). Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry*, **35**, 691–697.
79. Dill, K. A. & Bromberg, S. (2003). Polymer elasticity and collapse. In *Molecular Driving Forces*, chapt. 32, pp. 621–626, Garland Science, New York.
80. Gillespie, B. & Plaxco, K. W. (2004). Using protein folding rates to test protein folding theories. *Annu. Rev. Biochem.* **73**, 837–859.
81. Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Protein Struct. Funct. Genet.* **6**, 87–103.
82. Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Struct. Biol.* **4**, 630–634.
83. Chamberlain, A. K. & Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry*, **37**, 1736–1742.
84. Mayor, U., Grossmann, J. G., Foster, N. W., Freund, S. M. V. & Fersht, A. R. (2003). The denatured state of Engrailed Homeodomain under denaturing and native conditions. *J. Mol. Biol.* **333**, 977–991.
85. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.
86. Rami, B. R. & Udgaonkar, J. B. (2002). Mechanism of formation of a productive molten globule form of barstar. *Biochemistry*, **41**, 1710–1716.
87. Rami, B. R., Krishnamoorthy, G. & Udgaonkar, J. B. (2003). Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar. *Biochemistry*, **42**, 7986–4000.
88. Shimada, J. & Shakhnovich, E. I. (2002). The ensemble kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl Acad. Sci. USA*, **99**, 11175–11180.
89. Lyubovitsky, J. G., Gray, H. B. & Winkler, J. R. (2002). Mapping the cytochrome c folding landscape. *J. Am. Chem. Soc.* **124**, 5481–5485.
90. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1985). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet.* **21**, 167–195.
91. Staniforth, R. A., Dean, J. L. E., Zhong, Q., Zerovniki, E., Clarke, A. R. & Waltho, J. P. (2000). The major transition state in folding need not involve the immobilization of side chains. *Proc. Natl Acad. Sci. USA*, **97**, 5790–5795.
92. Pabit, S. A., Roder, H. & Hagen, S. J. (2004). Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry*, **43**, 12532–12538.
93. Frauenfelder, H., Fenimore, P. W., Chen, G. & McMahon, B. H. (2006). Protein folding is slaved to solvent motions. *Proc. Natl Acad. Sci. USA*, **103**, 15469–15472.
94. Chan, H. S. & Dill, K. A. (1990). Origin of structure in globular proteins. *Proc. Natl Acad. Sci. USA*, **87**, 6388–6392.
95. Hummer, G., Garcia, A. E. & Garde, S. (2001). Helix nucleation kinetics from molecular simulations in explicit solvent. *Proteins: Struct. Funct. Genet.* **42**, 77–84.
96. Garcia-Mira, M. M., Sadqui, M., Fischer, N., Sanchoz-Rulz, J. M. & Munoz, V. (2002). Experimental identification of downhill protein folding. *Science*, **298**, 1291–1295.
97. Yang, W. Y., Pitera, J. W., Swope, W. C. & Gruebele, M. (2004). Heterogeneous folding of the trpzip hairpin: full atom simulation and experiment. *J. Mol. Biol.* **336**, 241–251.
98. Shastry, M. C. R., Agashe, V. R. & Udgaonkar, J. B. (1994). Quantitative analysis of the kinetics of denaturation and renaturation of barstar in the folding transition zone. *Protein Sci.* **3**, 1409–1417.
99. Nolting, B., Golbik, R. & Fersht, A. R. (1995). Submillisecond events in protein folding. *Proc. Natl Acad. Sci. USA*, **92**, 10668–10672.
100. Lazaridis, T. & Karplus, M. (1997). "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science*, **278**, 1928–1931.
101. Ozkan, S. B., Dill, K. A. & Bahar, I. (2002). Fast folding kinetics, hidden intermediates, and the sequential stabilization model. *Protein Sci.* **11**, 1958–1970.
102. Bhuyan, A. K. & Udgaonkar, J. B. (1999). Observation of multistate kinetics during the slow folding and unfolding of barstar. *Biochemistry*, **38**, 9158–9168.
103. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.
104. Lu, J. & Dahlquist, F. W. (1992). Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry*, **31**, 4749–4756.
105. Nishimura, C., Dyson, H. J. & Wright, P. E. (2002). The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate. *J. Mol. Biol.* **322**, 483–489.
106. Agashe, V. R. & Udgaonkar, J. B. (1995). Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry*, **34**, 3286–3299.
107. Lubienski, M. J., Bycroft, M., Freund, S. M. V. & Fersht, A. R. (1994). Three-dimensional solution structure and 13C assignments of barstar using nuclear magnetic resonance spectroscopy. *Biochemistry*, **33**, 8866–8877.

Edited by K. Kuwajima

(Received 16 February 2007; received in revised form 10 April 2007; accepted 22 April 2007)
Available online 27 April 2007