doi:10.1016/j.jmb.2010.08.046 J. Mol. Biol. (2010) 403, 430–445

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

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

Evidence for Initial Non-specific Polypeptide Chain Collapse During the Refolding of the SH3 Domain of PI3 Kinase

Amrita Dasgupta and Jayant B. Udgaonkar*

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

Received 10 May 2010;
received in revised form
22 August 2010;
accepted 24 August 2010
Available online
15 September 2010

Edited by K. Kuwajima

Keywords:
PI3K SH3 domain;
burst-phase;
collapse;
continuous;
FRET

Refolding of the SH3 domain of PI3 kinase from the guanidine hydrochloride (GdnHCl)-unfolded state has been probed with millisecond (stopped flow) and sub-millisecond (continuous flow) measurements of the change in fluorescence, circular dichroism, ANS fluorescence and three-site fluorescence resonance energy transfer (FRET) efficiency. Fluorescence measurements are unable to detect structural changes preceding the rate-limiting step of folding, whereas measurements of changes in ANS fluorescence and FRET efficiency indicate that polypeptide chain collapse precedes the major structural transition. The initial chain collapse reaction is complete within 150 μs. The collapsed form at this time possesses hydrophobic clusters to which ANS binds. Each of the three measured intra-molecular distances has contracted to an extent predicted by the dependence of the FRET signal in completely unfolded protein on denaturant concentration, indicating that contraction is non-specific. The extent of contraction of each intra-molecular distance in the collapsed product of sub-millisecond folding increases continuously with a decrease in [GdnHCl]. The gradual contraction is continuous with the gradual contraction seen in completely unfolded protein, and its dependence on [GdnHCl] is not indicative of an all-or-none collapse reaction. The dependence of the extent of contraction on [GdnHCl] was similar for the three distances, indicating that chain collapse occurs in a synchronous manner across different segments of the polypeptide chain. The sub-millisecond measurements of folding in GdnHCl were unable to determine whether hydrophobic cluster formation, probed by ANS fluorescence measurement, precedes chain contraction probed by FRET. To determine whether hydrogen bonding plays a role in initial chain collapse, folding was initiated by dilution of the urea-unfolded state. The extent of contraction of at least one intra-molecular distance in the collapsed product of sub-millisecond folding in urea is similar to that seen in GdnHCl, and the initial contraction in urea too appears to be gradual.

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

Abbreviations used: FRET, fluorescence resonance energy transfer; PI3K SH3, SH3 domain of PI3 kinase; GdnHCl, guanidine hydrochloride; SAXS, small-angle X-ray scattering; ANS, 1-anilino-naphthalene-8-sulfonate; TNB, thionitrobenzoate; DTNB, 5, 5′-dithiobis (2-nitrobenzoic acid).

Introduction

Proteins are evolutionarily selected heteropolymers, but their response to solvent change appears to be very similar to that of simple homopolymers.[1,2] In particular, unfolded protein chains undergo global contraction when transferred from a good to

a bad solvent.[3,4] This contraction has been captured by equilibrium fluorescence resonance energy transfer (FRET) studies, in ensemble[5–9] as well as single molecule[10–12] measurements, which show that unfolded protein chains contract upon denaturant dilution. Kinetic FRET experiments reveal the diffusive and gradual nature of the initial chain contraction in the sub-millisecond time domain, upon transfer of unfolded protein from unfolding to refolding conditions.[6,13–15] The polymer-like dynamics of protein molecules is not restricted to unstructured polypeptide chains. For example, the diffusive, gradual swelling of the protein chain during the unfolding of monellin, which happens concomitantly with the loss of specific structure, is describable by a simple model of polymer physics.[16] But the role of polymer-like dynamics in the initiation as well as the progression of structure formation during protein folding is still not fully understood.

Polymer chains contract when intra-chain interactions dominate over chain–solvent interactions, and the consequent chain collapse reduces the conformational space that needs to be sampled by the diffusing protein chain in its search for the native conformation. It is commonly assumed that polypeptide chain collapse during protein folding is driven by non-specific intra-chain hydrophobic interactions.[3,17] Because of its non-specific nature, hydrophobic collapse is expected to be very rapid, leading to the formation of loose clusters of non-polar side chains, which remain hydrated, in a fluid-like contact with each other.[18] But it is also possible that chain collapse might, instead, be initiated by the formation of backbone hydrogen bonds, when intra-chain hydrogen bonding becomes progressively stronger at lower concentrations of denaturant,[19,20] resulting in intra-chain hydrogen bonds replacing the protein–denaturant hydrogen bonds that dominate at high concentrations of denaturant.[21] Such intra-chain hydrogen bonds could form in a non-specific, long-range manner,[14,22–24] leading to non-specific chain collapse, or in a specific, short-range manner leading to the formation of secondary structures that are compact and facilitate interactions between non-polar side chains.[25,26] A recurring question regarding the initiation of the protein folding reaction concerns the interplay between hydrophobic collapse and the formation of specific structure.[13,14,27–30]

For folding reactions, heteropolymer theory predicts that chain collapse should occur before the formation of native structure.[31] Direct measurements of chain collapse during folding have indicated that collapse precedes any significant structure formation for some,[3,32–34] but not all[35–37] proteins. Ultrafast measurements of the initial chain collapse reactions of several proteins[38–42] suggest that collapse occurs in the ~10 μs or faster time domain, and that structure forms in the ~100 μs and slower time domain. For proteins that complete folding within a few microseconds, it is not possible to resolve, temporally, chain collapse from structure formation,[43] but it appears that ultrafast folding could be the consequence of a pre-existing chain collapse in the unfolded state.[44] The extent of chain collapse might itself be modulated by long-range, intra-chain electrostatic interactions.[17,45,46] Much remains to be understood about the specificity of chain collapse that occurs initially during folding, and about what drives it.

SH3 domains are small, all-β proteins. A typical member of this family is the SH3 domain of PI3 kinase (PI3K SH3), which has 83 residues threaded through five β-strands and two helix-like turns (Fig. 1).[47] Like many other small proteins, SH3 domains had appeared to be two-state folders, with intermediates not being populated significantly on their folding pathways.[48,49]. But sparsely populated intermediates have been identified in nuclear magnetic resonance (NMR) spectroscopy studies, as in the case of the Fyn SH3 domain.[50,51] A non-native helical folding intermediate[52,53] has been identified by circular dichroism (CD) and small-angle X-ray scattering (SAXS) studies of the folding of the src SH3 domain at low temperature (where hydrophobic interactions are weakened). Very recently, a partially folded intermediate, which forms at a late stage during the folding of the PI3K SH3 domain, has been identified.[54] Studies of the unfolding of the PI3K SH3 domain, using hydrogen exchange detected by mass spectrometry,[55] in the absence and in low concentrations of denaturant also suggested that the dissolution of structure during unfolding occurs in many steps; hence, refolding too must occur in many steps in the absence and in low concentrations of denaturant. Nevertheless, earlier refolding studies were not able to detect any intermediate on the refolding pathway, probably because a suitable high-resolution probe was not used.[48] The PI3K SH3 domain therefore presents itself as an attractive model system for using site-specific probes to determine whether folding begins by initial chain collapse, and to understand the interplay of chain collapse and structure formation.

In this study, the refolding of the PI3K SH3 domain was investigated in the presence of guanidine hydrochloride (GdnHCl), as well as urea, using multiple probes, including 1-anilino-naphthalene-8-sulfonate (ANS) fluorescence and steady-state multi-site FRET. Both stopped flow and continuous flow measurements were made with mixing dead times of 1.8 ms and 0.15 ms, respectively. It was seen that ANS binds, albeit weakly, to the refolding protein within 0.15 ms, indicating that the unfolded protein has transformed into a form containing solvent-accessible hydrophobic clusters competent

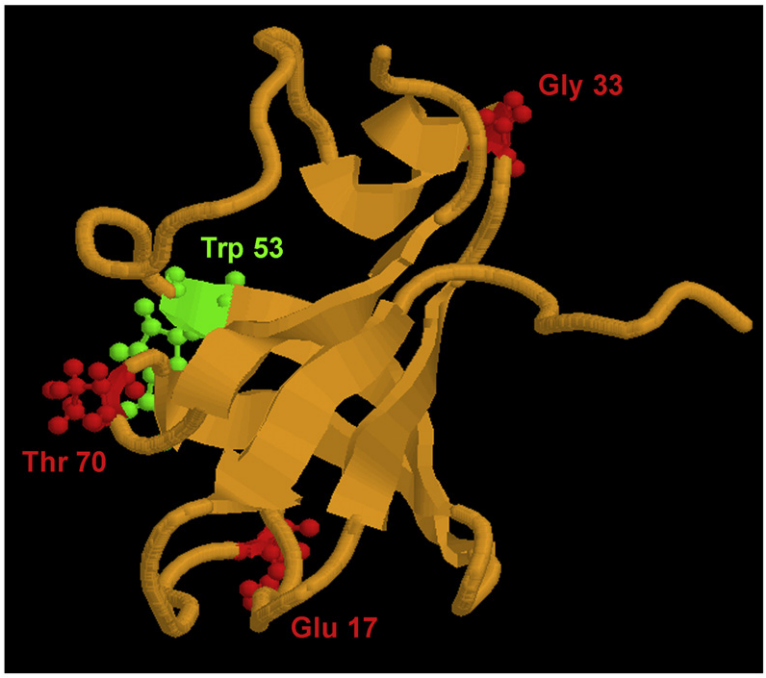

Fig. 1. Structure of the PI3K SH3 domain. The positions of W53 (green), E17 (red), G33 (red) and T70 (red) are shown. E17, G33 and T70 were mutated independently to C, to yield three different single cysteine-containing mutant proteins, Cys17, Cys33 and Cys70, respectively. The ribbon diagram was generated from PDB file 1pnj using Rasmol.

to bind ANS. Three-site FRET measurements indicated that the form of the protein bound was compact, and that the transition from the unfolded state to this compact state was gradual and non-specific. It appears that folding of the PI3K SH3 domain begins with a non-specific collapse of the polypeptide chain.

Results

Spectroscopic and kinetic characterization of the PI3K SH3 domain by fluorescence and CD measurements

The PI3K SH3 domain has a single tryptophan and seven tyrosine residues (Fig. 1). Fluorescence and CD measurements were done (Supplementary Data Figs S1 and S2) to show that the protein used in this study is identical in its spectroscopic properties, as well as in its stability and folding kinetics, to that used in earlier studies.[48,54,56]

Millisecond measurements of ANS fluorescence-monitored kinetics

The hydrophobic dye ANS binds to solvent-exposed hydrophobic clusters on proteins.[57] This binding results in an increase in fluorescence quantum yield of the dye and a blue shift in the fluorescence spectrum. In most instances, it was found that ANS does not bind to the native or the unfolded forms of proteins, including the PI3K SH3 domain.[56]

Figure 2a shows that when refolding is carried out in the presence of ANS, there is a burst phase increase in ANS fluorescence, within the dead time of stopped-flow mixing (2 ms). This initial jump in ANS fluorescence is relatively small, and corresponds to an increase of only 20 – 30 % of the ANS fluorescence of unfolded protein. Subsequently, there is a decrease in ANS fluorescence as the protein completes folding, and this decrease occurs in two kinetic phases for folding in concentrations of GdnHCl <0.8 M. The amplitude of the burst phase increase in fluorescence, as well as the observed rate constant for the fast phase of the decrease in ANS fluorescence, decreased with an increase in [GdnHCl]. The inset in Fig. 2a shows that the amplitude of the burst phase increase in ANS fluorescence becomes smaller with an increase in [GdnHCl], even at concentrations of GdnHCl that correspond to the native protein baseline of the equilibrium unfolding curve. The observed refolding rate constants of the decrease in ANS fluorescence match those measured using intrinsic tyrosine fluorescence as the probe (Fig. 2b). Separate experiments (data not shown) indicated that the presence of ANS in the concentration range 100 – 540 μM did not alter the observed refolding rate constants.

Thionitrobenzoate (TNB) quenches the fluorescence of W53 by FRET

The use of multi-site FRET to measure the changes in molecular dimensions, which occur during folding or unfolding reactions, is well documented.[6,8,13,16,32,58–60] The sole tryptophan

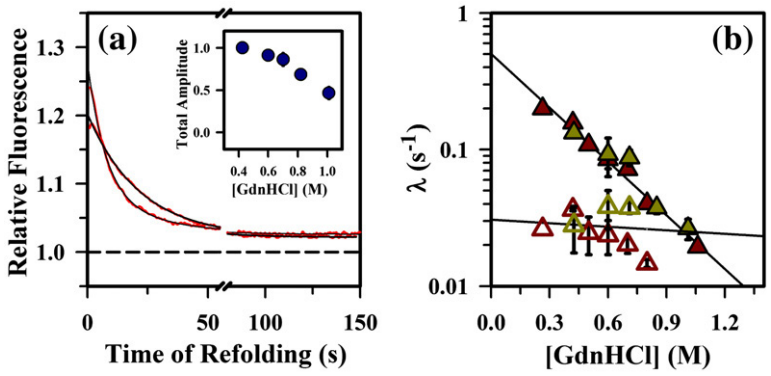

Fig. 2. ANS fluorescence-monitored kinetics of the PI3K SH3 domain at pH 7.2, 25 °C. (a) Representative kinetic traces of refolding in 0.4 M and 0.8 M GdnHCl (top to bottom) in the presence of ANS, starting from unfolded protein in 4 M GdnHCl. The kinetics of refolding was monitored by measurement of the total ANS fluorescence above 450 nm. The continuous black lines through the data show fits to a single-exponential equation for the kinetic trace obtained in 0.8 M GdnHCl, and to a double-exponential equation for the trace in 0.4 M GdnHCl. The broken line represents the ANS fluorescence of unfolded protein in 4 M GdnHCl. The inset shows the dependence of the total amplitude (●) on the concentration of GdnHCl present during refolding. The amplitudes are normalized to a value of 1 for the amplitude of the ANS fluorescence change observed for the refolding in 0.4 M GdnHCl. (b) A comparison of the fast and slow observed refolding rate constants monitored by intrinsic fluorescence at 300 nm (▲ and Δ) and ANS fluorescence above 450 nm (▲ and Δ). The continuous lines through the refolding constant data are least-squares fits to $\log\lambda = \log\lambda_{water} + m_f[\text{GdnHCl}]$. The values of the constants for both phases are similar to those shown in the legend to Supplementary Data Fig. S1. The error bars represent standard deviations obtained from three independent experiments.

residue in the sequence was chosen to serve as the donor for FRET measurements of the changes in intra-molecular distances that occur during the folding of the PI3K SH3 domain. The small TNB moiety was chosen to serve as the FRET acceptor.[5,61] Since the wild type protein has no cysteine residue that the TNB moiety could be attached to, three single cysteine-containing mutant proteins (Cys17, Cys33 and Cys70) were made by site-directed mutagenesis, and the TNB moiety was attached to the cysteine thiol by reacting each protein with 5, 5′-dithiobis (2-nitrobenzoic acid) (DTNB) (see Materials and Methods).

Figure 3a–c show the fluorescence spectra of Cys17, Cys17-TNB, Cys33, Cys33-TNB, Cys70 and Cys70-TNB in the folded and the unfolded forms. It is seen that the TNB adduct quenches the fluorescence of W53 in the native as well as the unfolded proteins, that for each pair the extent of quenching is greater in the native state than in the unfolded state, and that the extent of quenching in the native state depends on the position of the TNB adduct. The quenching of W53 fluorescence due to FRET is seen to depend on the separation between the donor and the acceptor in the protein. It is also seen that the difference between the fluorescence signals of the native and unfolded protein is small for all three unlabeled proteins: for each unlabeled protein, the quantum yield of the donor (W53) does not change by more than 12 %, as determined from the areas under the spectra of the native and unfolded proteins. Nevertheless, in subsequent FRET experiments, fluorescence was monitored at 387 nm to ensure that the fluorescence of the donor does not change significantly on refolding.

The insets in Fig. 3a–c show that when intrinsic fluorescence is used to monitor unfolding, the equilibrium unfolding curves of the labeled and the unlabeled proteins overlap well for each single cysteine-containing mutant protein. Hence, the stability of each mutant variant is unaffected by the presence of the TNB adduct. The stability of all three proteins, labeled as well as unlabeled, is very similar.

Refolding kinetics of the unlabeled and the TNB-labeled proteins in GdnHCl

The refolding kinetics of each unlabeled and labeled mutant protein was measured using fluorescence as the probe. For the labeled proteins, fluorescence was measured at 387 nm. At this wavelength, the fluorescence of the donor in the unlabeled protein does not change (see above); hence, the change in the fluorescence of the donor in the labeled protein represents the change in the FRET signal.[60] For the unlabeled proteins, fluorescence was measured at 300 nm. Figure 3d–f show that for each pair of labeled and unlabeled protein, the fast as well as the slow refolding rate constants are unaffected by the presence of the TNB adduct. Moreover, the observed rate constants and their dependences on [GdnHCl] are very similar for the different labeled and unlabeled mutant proteins, and are very similar to those of the wild type protein.

Figure 4a–c compare the equilibrium amplitudes of the fluorescence change at 387 nm to the kinetic amplitudes, for each pair of unlabeled and labeled protein, Cys17 and Cys17-TNB, Cys33 and Cys33-TNB, and Cys70 and Cys70-TNB, respectively. The

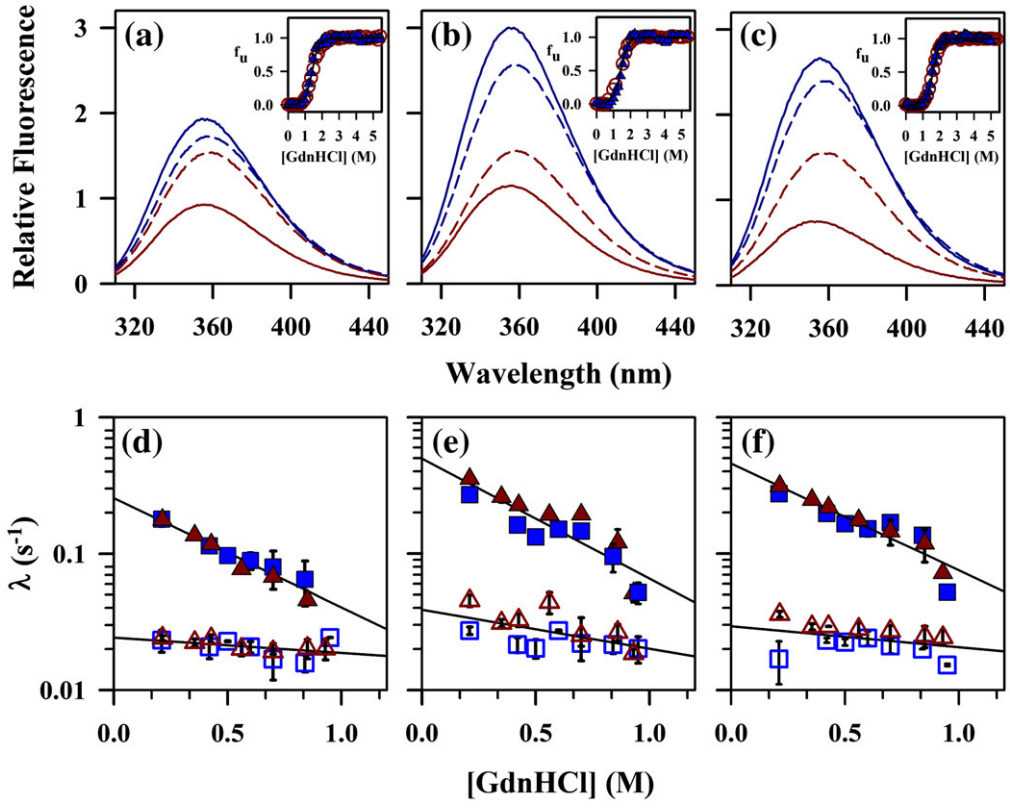

Fig. 3. Spectroscopic properties, stabilities and refolding kinetics of the unlabeled and labeled single cysteine-containing mutant variants of the PI3K SH3 domain. Fluorescence emission spectra of the unlabeled (blue lines) and labeled (red lines) proteins in their native (continuous lines) and unfolded (broken lines) states are shown for (a) Cys17 and Cys17-TNB, (b) Cys33 and Cys33-TNB, and (c) Cys70 and Cys70-TNB. Fluorescence was excited at 295 nm. In each panel, the data were normalized to a value of 1 for the fluorescence at 387 nm of labeled unfolded protein in 4 M GdnHCl. The insets show the equilibrium unfolding curves of the corresponding unlabeled (blue symbols) and labeled (red symbols) proteins. The data for each pair of labeled and unlabeled proteins were fit together to a two-state model, and the fits yielded values for C_m and ΔG_u of 1.4 M and 4.1 kcal/mol (Cys17 and Cys17-TNB), 1.4 M and 4.2 kcal/mol (Cys33 and Cys33-TNB) and 1.5 M and 4.7 kcal/mol (Cys70 and Cys70-TNB). (d–f) The dependence on GdnHCl concentration of the observed rate constants for the fast and slow phases of refolding are shown for (d) Cys17 and Cys17-TNB, (e) Cys33 and Cys33-TNB, and (f) Cys70 and Cys70-TNB. The filled and open symbols represent the observed rate constants of the fast and slow refolding phases, respectively. For monitoring folding of the labeled proteins (red symbols), fluorescence was excited at 295 nm and measured at 387 nm; while for the unlabeled proteins (blue symbols), fluorescence was excited at 268 nm and monitored at 300 nm. The continuous lines through the refolding constant data are least-squares fits to the equation: $\log\lambda = \log\lambda_{water} + m_f[\text{GdnHCl}]$ and yielded values for λ^f_{water} and m^f_f of 0.22 s^{-1} and –0.23 M^{-1} (Cys17 and Cys17-TNB), 0.4 s^{-1} and –0.3 M^{-1} (Cys33 and Cys33-TNB) and 0.4 s^{-1} and –0.3 M^{-1} (Cys70 and Cys70-TNB). The values for λ^s_{water} and m^s_f obtained were 0.02 s^{-1} and –0.01 M^{-1} (Cys17 and Cys17-TNB), 0.03 s^{-1} and –0.017 M^{-1} (Cys33 and Cys33-TNB) and 0.02 s^{-1} and –0.02 M^{-1} (Cys70 and Cys70-TNB). The error bars shown for the data in d–f represent the spread in the data from two separate experiments.

equilibrium unfolding transitions of Cys17, Cys33 and Cys70 show that the increase in fluorescence at 387 nm is small, and that the increase is comparable in magnitude to that seen for *N*-acetyl-L-tryptophanamide with an increase in denaturant concentration.[6] Nevertheless, the equilibrium unfolding transitions appear to retain an essentially sigmoidal shape. For each unlabeled proteins, the time zero points of the kinetic traces appear to fall on the extrapolated unfolded protein baseline, within experimental error.

Clearly sigmoidal equilibrium unfolding transitions were seen for the labeled proteins. When refolding was commenced from 4 M GdnHCl, a substantial (20 – 35%) fraction of the total change in fluorescence signal occurred in a sub-millisecond burst phase, within the 2 ms dead time of stopped-flow mixing. The time zero points of the kinetic traces of refolding were found to fall on the extrapolated unfolded protein baseline, and the infinity points fall on the equilibrium unfolding transition.

Dependence of the FRET efficiency and the D–A distance on the concentration of GdnHCl

The fluorescence values of the time zero points of the kinetic traces of folding for the labeled and unlabeled proteins, as well as the fluorescence values for the unfolded protein in different high concentrations of GdnHCl, were translated into FRET efficiency using Eq. (1). FRET efficiency is plotted against [GdnHCl] for the W53-C17TNB distance in Fig. 4d, for the W53-C33TNB distance in Fig. 4e and for the W53-C70TNB distance in Fig. 4f. For each intra-molecular distance, it is seen that the FRET efficiency in the product of sub- millisecond folding decreases in an asymptotic manner. This decrease merges with that seen for the fully unfolded protein, and a final limiting value is reached at a high concentration of GdnHCl. The dependence of the FRET efficiency on [GdnHCl] does not extrapolate at zero denaturant to the FRET efficiency measured in the native protein. This indicates that all three intra-molecular distances measured in the product of sub-millisecond folding are significantly larger than they are in the fully folded protein.

The FRET efficiency data in Fig. 4d–f were converted into intra-molecular distance data using

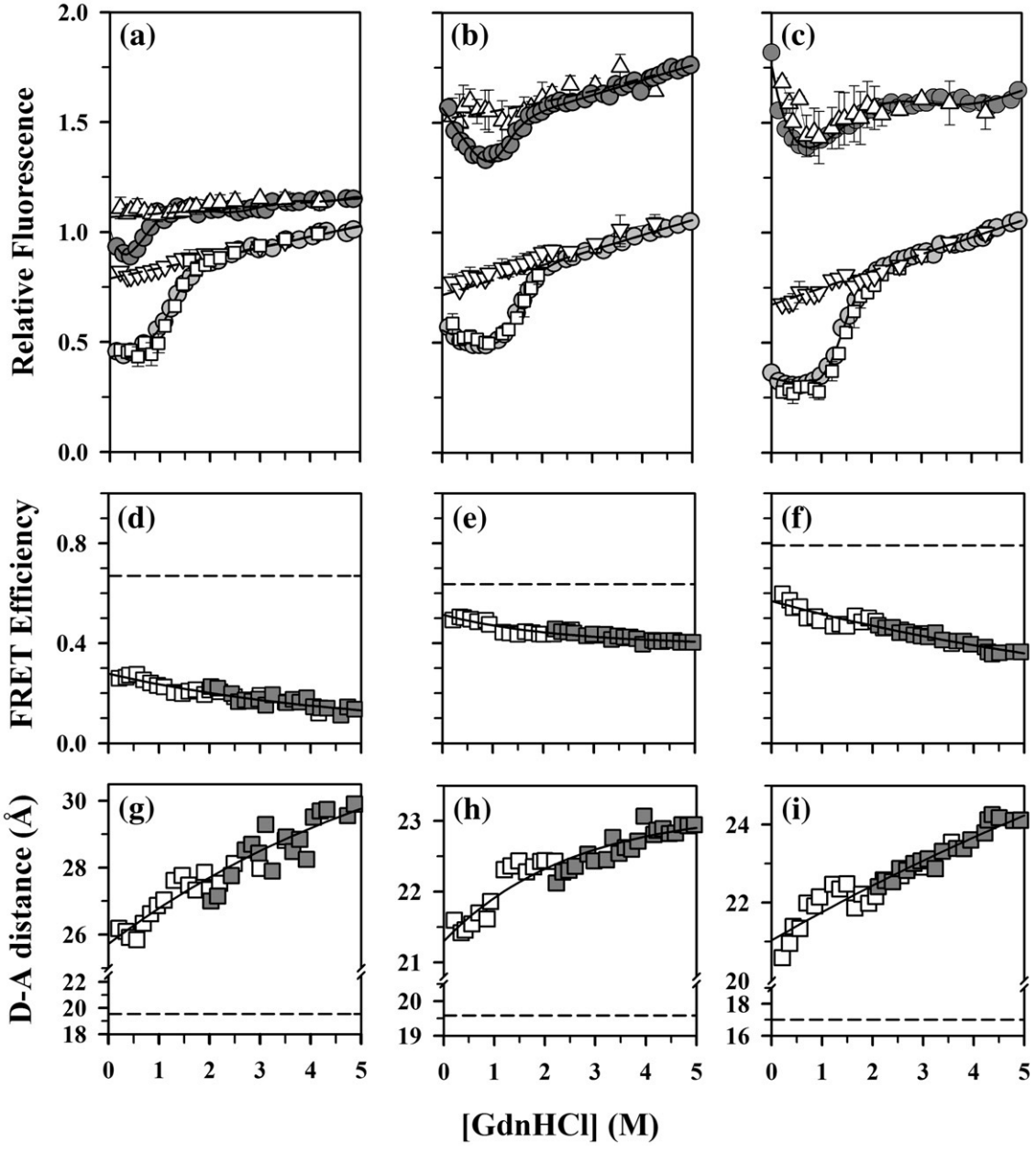

Fig. 4 (*legend on next page*)

Eq. (2). In order to do so, it was first necessary to determine the value of the Förster's distance R_0, which is given by Eq. (3), and the values of the overlap integral J and the quantum yield of the donor Q_D, were determined as described.[16] These values are summarized in Supplementary Data Table S1. The third parameter in Eq. (3) is the refractive index n and values of 1.333 for native protein in native buffer and 1.400 for unfolded protein in 4 M GdnHCl were used in Eq. (3). The donor, W53, is exposed to the solvent in the native state, and each cysteine residue is at a position where it is exposed to the solvent (Fig. 1). The absorbance spectra of the three labeled proteins confirmed that the TNB adducts are fully solvent-exposed (data not shown). Hence, a value of 2/3 was used for κ^2 in Eq. (3).[6,32,60] Table 1 shows that the values of R_0 determined for all three FRET pairs are nearly identical for the native and unfolded proteins. Hence, the value of R_0 for the burst phase product of stopped-flow mixing (the product of sub-millisecond folding) was assumed to be identical with that of the unfolded protein.

Figure 4g–i show how the unfolded form and the burst phase product of stopped-flow mixing contract upon a decrease of [GdnHCl], for Cys17-TNB, Cys33-TNB and Cys70-TNB, respectively. It is seen that for all three intra-molecular distances, the contraction has a continuous and gradual dependence on the denaturant concentration. This dependence appears to be describable by a single-exponential equation.

Figure 5a shows how the fractional change in FRET efficiency in the burst phase product of stopped-flow mixing and in the unfolded state, which was determined using Eq. (4), depend on [GdnHCl]. It is seen that the dependence values for the three FRET pairs overlap well within experimental error. Similarly, the dependences on [GdnHCl] of the fractional contraction of the three intra-molecular distances, shown in Fig. 5b, which were determined using Eq. (5), overlap for the three distances. These results suggest that the three intra-molecular distances in the product of sub-millisecond folding (the burst phase product of stopped-flow mixing) contract in a synchronous manner upon a decrease of [GdnHCl].

Sub-millisecond measurements of chain collapse using a continuous-flow mixer

Both the ANS fluorescence signal and the FRET signal change within the 6 ms dead time of stopped-flow mixing. Although the amplitudes of these signal changes that accompany the formation of the burst phase product of stopped flow mixing were easily determined (Figs. 2 and 4), it was important to try to determine the rate constant of formation of this burst phase product. In order to do so, a continuous mixing apparatus with a mixing dead time of 0.15 ms was used.[14] Figure 6a and b merge data of continuous-flow mixing and stopped-flow mixing for ANS fluorescence-monitored refolding of the wild type PI3K SH3 domain in 0.4 M and 0.7 M GdnHCl, respectively. The data merge well, but no kinetic process can be discerned in the first millisecond of folding. It appears that the initial jump in ANS fluorescence has occurred within the 0.15 ms dead time of continuous-flow mixing. Similarly, when the folding of Cys17-TNB (Fig. 6c) or of Cys70-TNB (Fig. 6d) was monitored by FRET in 0.4 M and 1.0 M GdnHCl, no kinetic process was discerned during the first millisecond of folding. Again, the continuous mixing kinetic trace for the

Fig. 4. Kinetic *versus* equilibrium amplitudes of folding monitored by FRET. (a) Cys17 and Cys17-TNB, (b) Cys33 and Cys33-TNB and (c) Cys70 and Cys70-TNB. FRET was measured by measurement of the change in fluorescence at 387 nm. (a–c) Symbols (●) and (●) represent the equilibrium unfolding transitions for the unlabeled and the labeled proteins, respectively; (△) and (▽) represent the time zero points of the kinetic traces of refolding for the unlabeled and the labeled proteins, respectively; (□) represents the infinity points of the kinetic traces of refolding of the labeled proteins. The broken lines in a–c represent the extrapolated unfolded protein baselines. The continuous lines through the equilibrium unfolding data of the labeled proteins represent fits to a two-state N↔U model for unfolding. The continuous lines for the unlabeled proteins are drawn by inspection only. (d–f) show the dependence of the FRET efficiency on GdnHCl concentration in the product of the sub-millisecond refolding reaction for Cys17-TNB, Cys33-TNB and Cys70-TNB, respectively. (■) The FRET efficiency (E) in the completely unfolded forms that are present in different concentrations of GdnHCl in the unfolded protein baseline of the equilibrium unfolding transition for the labeled proteins. (□) The FRET efficiency of the forms that are the products of the sub-millisecond phase of refolding. The continuous lines through the data represent fits to a single-exponential equation: $E = E(0) + a\exp^{-m[D]}$ where E (0) is the FRET efficiency in 0 M GdnHCl, a is the amplitude and m is the constant that defines the dependence of E on the denaturant concentration, [D]. The broken line represents the FRET efficiency in 0 M GdnHCl for fully folded protein. The error bars represent the spread in the data from two or more different sets of experiments. (g–i) The contraction of different intra-molecular distances in the product of the sub-millisecond refolding of the PI3K SH3 domain: (g) the W53-C17TNB distance in Cys17-TNB, (h) the W53-C33TNB distance in Cys33-TNB and (i) the W53-C70TNB distance in Cys70-TNB. (■) The D–A distances in the completely unfolded forms that are present at concentrations of GdnHCl corresponding to the unfolded protein baseline of the equilibrium unfolding transition for the labeled proteins. (□) The D-A distance of the collapsed species that is the product of the sub-millisecond phase of refolding. The continuous lines through the data are fits to a single-exponential equation. The broken lines represent the D–A distance for fully folded protein in 0 M GdnHCl.

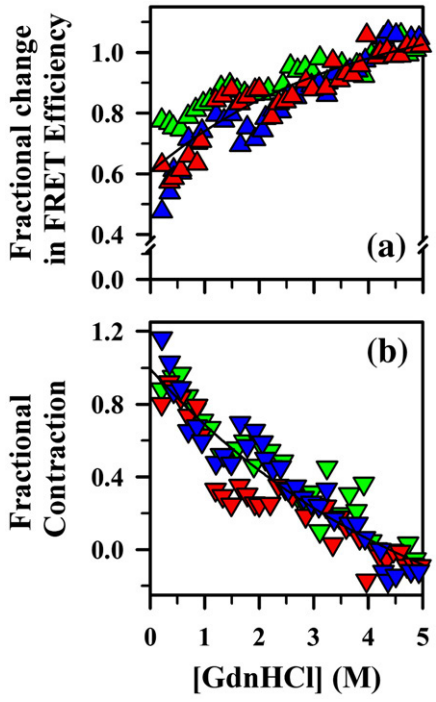

Fig. 5. Fractional contraction in the initially collapsed form. (a) The FRET efficiency data of Fig. 4d–f were normalized to the FRET efficiency in the folded and in the unfolded states using Eq. (4). (▲, ▲ and ▲) The fractional changes in FRET efficiency in different concentrations of GdnHCl for Cys17-TNB, Cys33-TNB and Cys70-TNB, respectively. The continuous line represents an exponential fit to the combined data in Fig. 4d–f. (b) The distance data of Fig. 4g–i were normalized to the distances in the unfolded protein and in the collapsed form in the absence of GdnHCl (obtained by extrapolation of the continuous lines in Fig. 4g–i to zero [GdnHCl]), using Eq. (5). (▼, ▼ and ▼) The fractional contraction in Cys17-TNB, Cys33-TNB and Cys70-TNB, respectively. The continuous line represents an exponential fit to the combined data in Fig. 4g–i.

first millisecond merges well with the slower stopped-flow kinetic trace, but it appears that the initial contraction of the protein upon folding from the unfolded state in 4 M GdnHCl to the product of sub-millisecond folding (the burst phase product of stopped-flow mixing) is complete within the 0.15 ms dead time of continuous-flow mixing. Hence, the initial collapse of the polypeptide chain during the folding of the SH3 domain is complete within 0.15 ms of the commencement of folding.

FRET measurement of initial chain collapse during folding in urea

To determine whether refolding in urea begins with chain collapse, the refolding of Cys70-TNB was monitored for three different concentrations of urea. The change in FRET was monitored by measurement of the change in the fluorescence signal at 387 nm (see above). Figure 7a compares the kinetic and equilibrium amplitudes for Cys70 and Cys70-TNB, and shows that a substantial change (up to 55 %) in the FRET signal occurs during the burst phase of stopped-flow mixing for folding in urea too. The data show a sigmoidal equilibrium unfolding transition for the labeled protein, and the mid-point is the same as that for the unlabeled protein. The time zero points of the kinetic traces of refolding were found to fall on the extrapolated unfolded protein baseline for Cys70 and Cys70-TNB, and the infinity points are seen to fall on the equilibrium unfolding transition for Cys70-TNB. It is seen that the FRET signal decreases continuously with an increase in [urea], both for the burst phase product of stopped-flow mixing and for the completely unfolded protein. Figure 7b shows the change in FRET efficiency with the urea concentration for the W53-C70TNB distance using Eq. (1). For this intra-molecular distance, it is seen that the FRET efficiency in the product of sub-millisecond folding decreases in an asymptotic manner. This decrease merges with that seen for the fully unfolded protein. The dependence of the FRET efficiency on [urea] does not extrapolate at zero denaturant to the FRET efficiency measured in the native protein. This indicates that the intra-molecular distance measured in the product of sub-millisecond folding is significantly larger than that in the fully folded protein.

Discussion

Intra-molecular distances in the unfolded state of the PI3K SH3 domain

In 4 M GdnHCl, the three intra-molecular distances measured by FRET for the unfolded protein are close in value (within ~2 Å) to the distances calculated for an unfolded polypeptide chain modeled as a random coil without accounting for the excluded volume effect.[62] All three distances expand when [GdnHCl] is increased from 4 M to 5 M (Fig. 4). Such continuous expansion of intra-molecular distances in the unfolded protein (Figs. 4 and 7) with an increase in denaturant concentration is the expected response of a simple polymer when the nature of the solvent is changed from bad to good.[63,64] Hence, it appears that the unfolded PI3K SH3 domain behaves like a random coil polymer, although it is not possible to completely rule out the existence of residual structure in the unfolded state.[65–67] The extent of expansion of the intra-molecular distances in the unfolded PI3K SH3 domain with an increase in [GdnHCl] is similar to the extent to which intra-molecular distances were

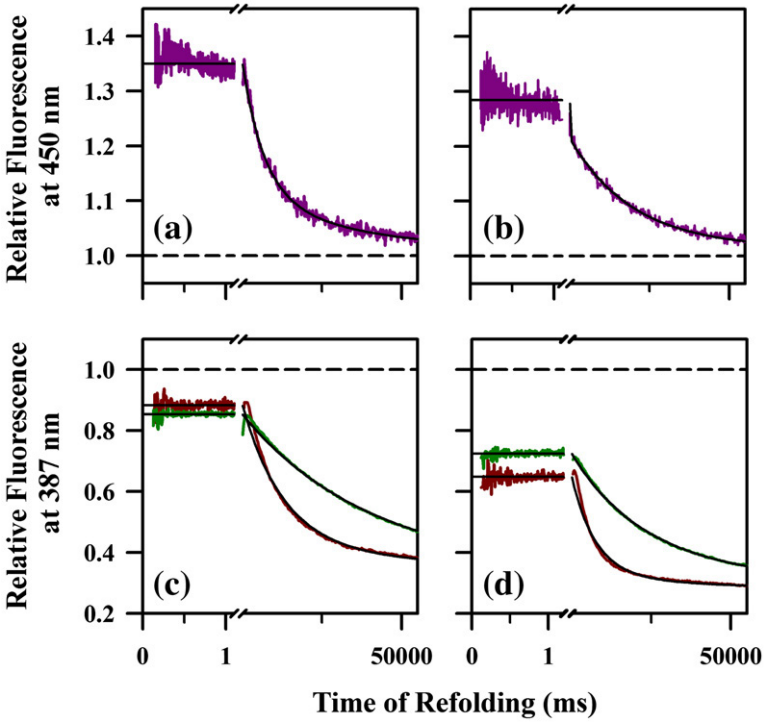

Fig. 6. Sub-millisecond refolding. Kinetic traces were measured using ANS fluorescence and FRET as probes in a continuous-flow mixer. (a and b) Sub-millisecond refolding kinetic traces at 0.4 M and at 0.7 M GdnHCl, respectively, for the folding of the wild type PI3K SH3 domain, monitored by measurement of ANS fluorescence. (c and d) Sub-millisecond refolding kinetic traces monitored by FRET in 0.4 M GdnHCl (red) and 1.0 M GdnHCl (green) for Cys17-TNB and Cys70-TNB, respectively. The initial 55 s of the kinetic traces obtained after stopped-flow mixing are also shown. (a–d) The data are normalized to a value of 1 for the unfolded protein in 4 M GdnHCl, represented by the broken line. The sub-millisecond data from the continuous-flow mixer and the millisecond data from the stopped-flow are normalized to each other as described in Materials and Methods.

observed to expand in unfolded barstar upon an increase in denaturant concentration.[5–7] Unfolded state expansion with an increase in denaturant concentration has been observed for several other proteins,[8–12] but not for all of the proteins that have been studied.[35,37]

Polypeptide chain collapse precedes the rate-limiting step during the folding of the PI3K SH3 domain

SH3 domains have served as model proteins for folding studies for many years. In particular, in the guise of two-state folders, they have provided very useful information on the transition state of folding.[68,69] It was reported for several other two-state folders, that polypeptide chain collapse is concomitant with structure formation during folding.[35,36] In these cases, SAXS measurements indicated that chain contraction occurs not initially in the sub-millisecond time domain but, surprisingly, in the millisecond – second time domain along with structure formation.[35–37] For one of these proteins, protein L, single-molecule FRET experiments have, however, indicated that chain contraction does indeed occur.[64,70] It has been suggested that SAXS measurements might not be as sensitive as FRET measurements for measuring chain contraction when folding reactions are initiated by denaturant dilution,[71] but this is certainly not always the case.[27,52,72] In the case of tendamistat, the absence of an initial collapse reaction was deduced from the absence of sub-millisecond change in intrinsic tryptophan fluorescence,[73] but several other proteins that have been shown by FRET to undergo sub-millisecond collapse reactions do not display sub-millisecond changes in intrinsic tryptophan fluorescence.[3,41,74,75] It now appears that even some proteins that once seemed very well approximated as two state folders might fold *via* well-populated collapsed states; collapse need not be concomitant with folding itself. [32,76]

In this study, it was shown that the PI3K SH3 domain undergoes chain collapse before the rate-limiting step in folding. Two probes, ANS fluorescence and FRET, were used to confirm this result; however, the rate of the collapse process could not be determined. Both probes of collapse, FRET as well as ANS fluorescence, indicated that the collapse process was complete within 0.15 ms, suggesting a lower limit of 20,000 s^{-1} for the rate constant of the collapse reaction. Rapid initial polypeptide chain contraction consequent to solvent change has been reported for other proteins.[3,4,77] Initial chain contraction during refolding might occur within a few microseconds.[31,72,78]

What drives initial chain collapse during the folding of the PI3K SH3 domain?

It is difficult to dissect out the contributions of side chain chemistry and main chain hydrogen bonding to the overall process of compaction during folding[79] and particularly to the initial chain collapse

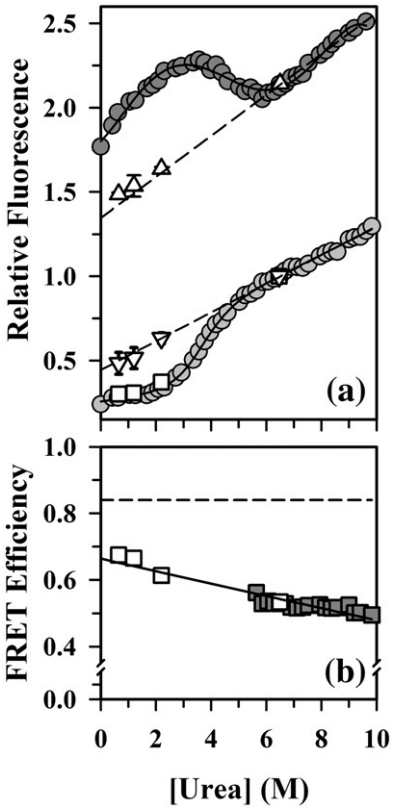

Fig. 7. Kinetic *versus* equilibrium amplitudes of the folding of Cys70 and Cys70-TNB in urea monitored by FRET. (a) (●) and (●) show the equilibrium unfolding transitions for Cys70 and Cys70-TNB, respectively; (Δ) and (▽) the time zero points of the kinetic traces of refolding for Cys70 and Cys70-TNB, respectively; (□) represents the infinity points of the kinetic traces of refolding for Cys70-TNB. The broken lines in (a) represent the extrapolated unfolded protein baselines for the labeled and the unlabeled proteins. The continuous line through the equilibrium unfolding data of the labeled protein represents a fit to a two-state N↔U model for unfolding. The continuous line through the equilibrium unfolding data of the unlabeled protein is drawn by inspection only. (b) The dependence of FRET efficiency on the concentration of urea in the product of the sub-millisecond refolding reaction for Cys70-TNB. (■) The FRET efficiency (E) in the completely unfolded forms that are present in different concentrations of GdnHCl in the unfolded protein baseline of the equilibrium unfolding transition for the labeled proteins. (□) The FRET efficiency of the forms that are the products of the sub-millisecond phase of refolding. The continuous lines through the data represent fits to a single-exponential equation: $E = E(0) + a\exp^{-m[D]}$ where E (0) is the FRET efficiency in the absence of GdnHCl, a is the amplitude and m is the constant that defines the dependence of E on the denaturant concentration, [D]. The broken line represents the FRET efficiency in the absence of GdnHCl for fully folded protein. The error bars represent the spread in the data from two independent sets of experiments.

reaction.[80] The observation that an initial jump in ANS fluorescence occurs within the first 0.15 ms of folding suggests that the initial collapse of the polypeptide chain might be driven by hydrophobic interactions between non-polar residues.[3,17,18,81] Such a hydrophobic collapse would result in the formation of solvated hydrophobic clusters to which ANS binds. But the FRET signal also changes within the first 0.15 ms, and it is not possible to temporally resolve the formation of solvent-accessible hydrophobic clusters measured by ANS fluorescence from the contraction of intra-molecular distances measured by FRET. It is therefore possible that the initial driving force for collapse is non-specific hydrogen bond formation between different segments of the polypeptide chain, and that such main chain hydrogen bond-driven collapse is followed by the formation of hydrophobic clusters, with both processes occurring within the first 0.15 ms.

In this study, the initial chain contraction was shown to occur during refolding in GdnHCl and in urea (Figs. 4 and 7). A recent study showed that upon dilution of a urea solution, used to initiate refolding, intra-chain hydrogen bonding in a protein becomes stronger while hydrophobic interactions remain unaffected.[20] The extent of chain contraction is seen to increase with a decrease in [urea] (Fig. 7) and it is possible that at least in the case of the refolding of the PI3K SH3 domain in urea, chain contraction might be driven by hydrophobic interactions and by intra-chain hydrogen bonding. It is possible that this is so for refolding in GdnHCl as well: the change in FRET efficiency that occurs upon the initial contraction of the measured intra-molecular distance is about the same in GdnHCl and in urea (Figs. 4 and 7). In this context, it has been shown that a model polypeptide devoid of any hydrophobic groups in its sequence undergoes a compaction upon dilution of GdnHCl.[23] This study, as well as other studies of the compaction of peptides deemed not to contain hydrophobic residues,[22,24] ascribes the compaction to the formation of non-specific intra-molecular hydrogen bonds. More work is needed to further delineate the contributions of hydrophobic interactions and hydrogen bonding in driving initial chain collapse during folding. It is noteworthy that in the case of the initial sub-millisecond folding reaction of barstar, FRET-monitored chain contraction was shown to occur faster than ANS fluorescence-monitored formation of hydrophobic clusters,[14] and it was suggested that the initial contraction might be driven by peptide hydrogen bond formation.

Nature of the initial chain collapse reaction during the folding of the PI3K SH3 domain

There has been considerable debate regarding the co-operativity of the initial collapse reaction during

protein folding.[6,13,14,30,71,82] The observation that all three intra-molecular distances in the burst phase (6 ms) product of stopped-flow mixing contract in a non-sigmoidal manner with a decrease of [GdnHCl] in which folding is commenced, and that each contraction in distance is continuous with that seen for fully unfolded protein (Fig. 4), suggests that the collapse reaction is not an all-or-none transition but is instead a continuous, gradual transition.[6,13] The monotonic burst phase decrease seen for each of the three intra-molecular distances in the PI3K SH3 domain is similar to that seen for each of 12 intra-molecular distances measured during the initial collapse reaction during the folding of barstar.[6,13] In the case of barstar, the gradual nature of the collapse reaction was apparent in an equilibrium model for the commencement of folding,[17] and was confirmed by sub-millisecond measurements of the rate constant of chain collapse: the rate constant was found to be independent of denaturant concentration.[14] For the PI3K SH3 domain, however, the collapse reaction is too fast for the rate constant to be determined using the continuous mixer with a mixing dead time of 0.15 ms; hence, it could not be confirmed by direct kinetic measurements that its collapse reaction is not an all-or-none cooperative transition.

The unfolding of the PI3K SH3 domain has been studied in the absence of denaturant as well as in the presence of low concentrations of denaturant (~0.5 M GdnHCl) by use of the hydrogen exchange detected by mass spectrometry. In these conditions, the unfolding reaction occurs in multiple steps.[55] Importantly, the final unfolding of an intermediate to the fully unfolded state does not occur in an all-or-none manner. These conditions are identical with the strongly stabilizing folding conditions utilized in that study, so the initial stage of refolding would be expected to be the reverse of the last stage of unfolding. Hence, it is likely that the initial collapse of the polypeptide chain during the folding of the PI3K SH3 domain is a gradual process.

The observation that the FRET signal of the product of the initial collapse reaction of the PI3K SH3 domain at any [GdnHCl] is predicted by a linear extrapolation of the unfolded protein FRET signal to that [GdnHCl],[6,13] suggests that the initial collapse reaction is non-specific. Such an inference is supported by the observation that the three intra-molecular distances contract to the same fractional extent (Fig. 5) with a decrease in denaturant concentration. In contrast, in the case of barstar, 12 intra-molecular distances in the product of sub-millisecond folding were observed to contract to different fractional extents that had different dependence on denaturant concentration, suggesting that chain collapse was not synchronous across different segments of the polypeptide chain.[13] Hence, initial polypeptide chain collapse during the folding of the PI3K SH3 domain appears to be synchronous, as would be expected for a chain collapse reaction that is completely non-specific and the consequence of a change from a good solvent to a bad solvent.[63,71] In this respect, the collapse reaction of the PI3K SH3 domain resembles the coil to globule transition of a simple polymer and the observation that the fractional contraction of the distance separating two positions along the sequence is not proportional to the number of residues separating the two positions is not what is expected for a simple coil to globule transition.[62,83,84]

Polypeptide collapse and secondary structure formation

The far-UV CD spectrum (Supplementary Data Fig. S2,) of the PI3K SH3 domain has an anomalous shape,[48,54] and the magnitude of the mean residue ellipticity is more typical of an unfolded protein than of a folded protein. Furthermore, the far-UV CD-monitored equilibrium unfolding curve of the PI3K SH3 domain has an unusual shape (Supplementary Data Fig. S2). The extrapolated unfolded protein baseline is relatively steep: the unfolded protein signal change for a 1 M change in denaturant concentration is only slightly less than the signal change in the transition zone for a similar change in denaturant concentration. Consequently, it intersects the native protein baseline at ~0.5 M GdnHCl. Equilibrium unfolding curve baselines that are steep in this manner, are more likely to reflect gradual structural folding and unfolding changes[5,8,85] rather than phenomenological dependence of the probe on the denaturant concentration, as is usually assumed. Whatever their origin might be, such baselines complicate the analysis of the change seen in the far-UV CD signal during refolding, especially because the absolute change in the signal is too small for millisecond measurements initiated by stopped-flow mixing.

Measurement of the change in the far-UV CD signal upon refolding was possible only over a narrow range of [GdnHCl] (0.9 – 1.3 M). The major fractional increase in far-UV CD occurs during the sub-second burst phase of manual mixing, and the minor observable kinetic phase accounts for the entire equilibrium amplitude of signal change when the latter is determined as the difference between the linearly extrapolated unfolded protein signal and the native protein signal (Supplementary Data Fig. S2). But the absolute increase in the far-UV CD signal even in the unobservable sub-second kinetic phase, is so small that it appears that the product of the initial collapse reaction has very little secondary structure in any case. In the case of barstar, the product of sub-millisecond folding does not possess secondary structure under marginally stabilizing conditions (1 M GdnHCl or 2.3 M urea), but does so

at lower denaturant concentrations.[3,45] In the case of the PI3K SH3 domain, the far-UV CD changes at the lower denaturant concentrations were too small to measure by stopped-flow mixing and too fast to measure by manual mixing. Hence, it could not be determined whether the product of sub-millisecond folding in strongly stabilizing conditions possesses secondary structure.

The initially collapsed form of the PI3K SH3 domain in marginally stabilizing conditions (~ 1 M GdnHCl) is compact, as measured by FRET, is solvated, as measured by its ability to bind to ANS, and has very little ellipticity at 222 nm. It was not possible to determine the secondary structural content of the initially collapsed form because the native state itself has a far-UV CD spectrum of very low intensity; consequently, the change in ellipticity during folding is very small (Supplementary Data Fig. S2). But it would not be unusual if the collapsed form does indeed have very little secondary structure: such collapsed forms have been seen for other proteins.[3,9,17,27,32,33,40,41,86,87] In future studies, it will be important to determine the type of interaction that would stabilize such a collapsed form. The interaction could be hydrophobic or non-specific hydrogen bonding interaction, could be short-range or long-range, and could be native or non-native. Understanding how a compact globule with very little, if any, specific structure is stabilized is crucial for understanding the subsequent development of specific structure during the folding of a protein. It is yet to be understood whether the folding reaction commencing *via* a non-specific collapse, preceding specific structure formation, is advantageous in any way, in comparison to it commencing *via* concomitant collapse and structure formation.[88–91] In order to understand properly the relevance of initial chain collapse to subsequent structure formation, current effort in the laboratory is focused on determining how the structure of the collapsed product can be tuned by a change in folding conditions.

Materials and Methods

Buffers and reagents

All reagents and buffers were of the highest purity obtainable from Sigma. Ultra-pure grade GdnHCl and urea were obtained from USB Corp. A 20 mM phosphate buffer (pH 7.2) was used for all experiments, which were all done at 25 °C. The concentrations of the stock solutions of GdnHCl and urea were determined by refractive index measurements on an Abbe refractometer. All experiments using the unlabeled single-cysteine mutants were done in the presence of 1 mM dithiothreitol (DTT) obtained from Invitrogen.

Protein expression, purification and labeling

The PI3K SH3 domain used in earlier studies had an N-terminal serine residue that was cleaved in 10–15 % of the protein molecules upon purification but this sequence heterogeneity did not affect the structure, stability or folding kinetics of the protein.[54] Nonetheless, it was decided to avoid the sequence heterogeneity altogether in this study. Hence, the DNA sequence of the clone was modified so that the expressed protein would not have an N-terminal serine residue. The sequence of the protein used in this study is:

AEGYQYRALYDYKKEREEDIDLHLGDILTVNKGSL-
VALGFSDGQEAKPEEIGWLNGYNETTGERGDFPG-
TYVEYIGRKKISP

The protein was purified as described,[54] and electrospray ionization mass spectrometry showed that the protein had the expected mass of 9276.2 Da. The purity of the protein was further confirmed by SDS-PAGE, which indicated that the protein was >98% pure.

The wild type PI3K SH3 domain contains one tryptophan at position 53 and has no cysteine residue. The mutant proteins Cys17, Cys33 and Cys70 each containing only a single cysteine residue at the position indicated in their names, were generated by single-site mutagenesis. They were sequenced and purified by the protocol used for the wild type protein.

The TNB-labeled proteins were obtained by incubating the protein in a 100-fold molar excess of DTNB in 6 M GdnHCl at pH 8.5. After incubation at room temperature for 2 h for completion of the reaction, the excess DTNB was removed using a desalting column (Sephadex-25, from Amersham Biosciences). Electrospray ionization mass spectrometry indicated that the proteins were labeled to >95%, with an increase in mass of 196 Da, as expected for the addition of the TNB group to the cysteine thiol.

Determination of protein concentrations

The concentrations of the wild type and unlabeled cysteine mutants were determined by measuring the absorbance at 280 nm. In the case of the TNB-labeled proteins, the TNB label contributes to the absorbance at 280 nm. Hence, a correction was applied to the measured absorbance readings as described.[6]

Equilibrium unfolding experiments

All experiments were done in a stopped-flow module (SFM-4, Biologic) in a fluorescence cuvette of 1 cm path-length. The wavelength for selective excitation of the tyrosines was 268 nm with a bandwidth of 4 nm, and fluorescence emission was measured through a 300 nm band-pass filter of bandwidth 10 nm (Asahi Spectra). The concentration of protein used in these experiments was 10 μM. For all FRET experiments, excitation of the tryptophan fluorescence was done at 295 nm with a bandwidth of 2 nm. Emission was monitored at 387 nm (with a bandwidth of 25 nm), for the unlabeled and the

labeled proteins. Far-UV CD experiments were carried out on a Jasco-J720 spectropolarimeter. The concentration of the protein used for experiments at 216 nm and 222 nm was 40 μM and 30 μM, respectively, and the path-length of the cuvette used was 0.2 cm.

Kinetic refolding experiments

Manual mixing experiments using far-UV CD

Far-UV CD-monitored refolding experiments were done by manually mixing unfolded protein in 4 M GdnHCl with native buffer at pH 7.2. The typical mixing dead time using a 0.2 cm cuvette was 9–10 s.

Stopped-flow mixing experiments

All tyrosine fluorescence, FRET and ANS fluorescence experiments were done in the SFM-4 stopped-flow module. Typically, a dead time of 2 ms was achieved using a cuvette of 0.08 cm path-length, 6.2 ms was achieved using a cuvette of 0.15 cm path-length and 10.8 ms was achieved using a cuvette of 0.2 cm path-length. The concentration used of the wild type and the unlabeled cysteine-containing mutant proteins was 20 μM for measurement of changes in tyrosine fluorescence. The concentration of the unlabeled proteins was 40 μM and that of the TNB-labeled proteins was 15 μM for experiments using FRET as the probe. The acquisition parameters were the same as those used for the equilibrium experiments. For experiments using ANS, the ANS concentration was 540 μM in the refolding buffer and the protein concentration was 45 μM. The wavelength for excitation was 295 nm with a bandwidth of 5 nm, and the emission wavelength was 450 nm with a bandwidth of 25 nm.

Continuous-flow mixing experiments

The design and operation of the mixer has been described.[14] The only difference between the previous and the current setup is that the 150 W Xe (Hg) lamp was replaced by a 500 W Xe (Hg) lamp. To determine the dead time of the instrument, the kinetics of the quenching of *N*-acetyl tryptophanamide fluorescence by *N*-bromosuccinamide were measured.[14] The current dead time of the instrument is 155 μs at a flow rate of 1 ml s^{-1}. All experiments using ANS fluorescence and FRET as probes were done with the settings described for the measurement of the kinetics using the stopped-flow module. The relative fluorescence intensity profile (S_{rel}) was obtained with respect to the fluorescence intensity profile of the unfolded protein in 4 M GdnHCl, using the relationship:

$$S_{rel} = (S_x - S_b) / (S_{up} - S_{ub})$$

where S_x is the fluorescence signal of the refolding protein, S_{up} is the fluorescence signal of the unfolded protein, and S_b and S_{ub} are the signals for the refolding buffer and unfolding buffer, respectively. For a direct comparison of the continuous-flow data with stopped-flow data, the latter were normalized to the signal of the unfolded protein in 4 M GdnHCl under identical conditions.

Data analysis

Determination of FRET efficiency

The efficiency of FRET, *E*, for any donor (D)–acceptor (A) pair is given by:

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

where, F_D is fluorescence intensity of the donor in the absence of the acceptor, and is measured in the unlabeled protein; F_{DA} is the fluorescence intensity of the donor in the presence of the acceptor, and is measured in the labeled protein.

E is related to the distance *R* separating donor and acceptor by:

$$E = \frac{R_0^6}{R_0^6 + R^6} \quad (2)$$

where R_o is the Förster's distance, which is given by:

$$R_0 = 0.211\left(Q_D J \kappa^2 n^{-4}\right)^{1/6} \quad (3)$$

In Eq. (3), Q_D is the quantum yield of the donor in the absence of the acceptor, *J* is the overlap integral, *n* is the refractive index, and κ^2 is the orientation factor. *J* and Q_D were determined as described.[6,16]

Calculation of the fractional change in FRET efficiency

The fractional change in FRET efficiency was calculated as:

$$F_E = \frac{(E_N - E_C)}{(E_N - E_U)} \quad (4)$$

where F_E is the fractional change in FRET efficiency, E_N is the FRET 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:

$$F_C = \frac{(D_U - D_C)}{(D_U - D_0)} \quad (5)$$

where F_C is the fractional contraction, D_U is the D–A distance in the unfolded state (in 4 M GdnHCl), 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 the absence of GdnHCl obtained by extrapolation.

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2010.08.046

Acknowledgements

We thank all the members of our laboratory for discussions. J.B.U. is a recipient of the JC Bose National Fellowship from the Government of India. This work was funded by the Tata Institute of

Fundamental Research, and by the Department of Science and Technology, Government of India.

References

1. de Gennes, P. G. (1985). Kinetics of collapse for a flexible coil. *J. Phys. Lett.* **46**, L639–L642.
2. Sanchez, I. C. (1979). Phase transition behavior of the isolated polymer chain. *Macromolecules*, **12**, 980–988.
3. Agashe, V. R., Shastry, M. C. & Udgaonkar, J. B. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature*, **377**, 754–757.
4. Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. (1997). Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl Acad. Sci. USA*, **94**, 8545–8550.
5. Lakshmikanth, G. S., Sridevi, K., Krishnamoorthy, G. & Udgaonkar, J. B. (2001). Structure is lost incrementally during the unfolding of barstar. *Nat. Struct. Biol.* **8**, 799–804.
6. 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.
7. Saxena, A. M., Udgaonkar, J. B. & Krishnamoorthy, G. (2006). Characterization of intra-molecular distances and site-specific dynamics in chemically unfolded barstar: evidence for denaturant-dependent non-random structure. *J. Mol. Biol.* **359**, 174–189.
8. Jha, S. K. & Udgaonkar, J. B. (2009). Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc. Natl Acad. Sci. USA*, **106**, 12289–12294.
9. Huang, F., Ying, L. & Fersht, A. R. (2009). Direct observation of barrier-limited folding of BBL by single-molecule fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA*, **106**, 16239–16244.
10. Schuler, B., Lipman, E. A. & Eaton, W. A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, **419**, 743–747.
11. 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.
12. 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 synchrotron radiation circular dichroism spectroscopy. *Proc. Natl Acad. Sci. USA*, **104**, 105–110.
13. Sinha, K. K. & Udgaonkar, J. B. (2007). Dissecting the non-specific and specific components of the initial folding reaction of barstar by multi-site FRET measurements. *J. Mol. Biol.* **370**, 385–405.
14. Sinha, K. K. & Udgaonkar, J. B. (2008). Barrierless evolution of structure during the submillisecond refolding reaction of a small protein. *Proc. Natl Acad. Sci. USA*, **105**, 7998–8003.
15. Pletneva, E. V., Gray, H. B. & Winkler, J. R. (2005). Snapshots of cytochrome *c* folding. *Proc. Natl Acad. Sci. USA*, **102**, 18397–18402.
16. Jha, S. K., Dhar, D., Krishnamoorthy, G. & Udgaonkar, J. B. (2009). Continuous dissolution of structure during the unfolding of a small protein. *Proc. Natl Acad. Sci. USA*, **106**, 11113–11118.
17. Rami, B. R. & Udgaonkar, J. B. (2002). Mechanism of formation of a productive molten globule form of barstar. *Biochemistry*, **41**, 1710–1716.
18. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.
19. Bolen, D. W. & Rose, G. D. (2008). Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **77**, 339–362.
20. Holthauzen, L. M., Rosgen, J. & Bolen, D. W. (2010). Hydrogen bonding progressively strengthens upon transfer of the protein urea-denatured state to water and protecting osmolytes. *Biochemistry*, **49**, 1310–1318.
21. Makhatadze, G. I. & Privalov, P. L. (1992). Protein interactions with urea and guanidinium chloride. A calorimetric study. *J. Mol. Biol.* **226**, 491–505.
22. Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R. & Pappu, R. V. (2006). Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl Acad. Sci. USA*, **103**, 16764–16769.
23. 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.
24. Daidone, I., Neuweiler, H., Doose, S., Sauer, M. & Smith, J. C. (2010). Hydrogen-bond driven loop-closure kinetics in unfolded polypeptide chains. *PLoS Comput. Biol.* **6**, e1000645.
25. Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry*, **24**, 1501–1509.
26. Rose, G. D., Fleming, P. J., Banavar, J. R. & Maritan, A. (2006). A backbone-based theory of protein folding. *Proc. Natl Acad. Sci. USA*, **103**, 16623–16633.
27. 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.
28. Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I. & Fujisawa, T. (2004). Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness. *Proc. Natl Acad. Sci. USA*, **101**, 1171–1176.
29. Orevi, T., Ben Ishay, E., Pirchi, M., Jacob, M. H., Amir, D. & Haas, E. (2009). Early closure of a long loop in the refolding of adenylate kinase: a possible key role of non-local interactions in the initial folding steps. *J. Mol. Biol.* **385**, 1230–1242.
30. Sinha, K. K. & Udgaonkar, J. B. (2009). Early events in protein folding. *Curr. Sci.* **96**, 1053–1070.
31. Thirumalai, D. (1995). From minimal models to real proteins: time scales for protein folding kinetics. *J. Phys. I (France)*, **5**, 1457–1467.
32. 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.
33. Ratner, V., Amir, D., Kahana, E. & Haas, E. (2005). Fast collapse but slow formation of secondary

structure elements in the refolding transition of *E. coli* adenylate kinase. *J. Mol. Biol.* **352**, 683–699.
34. DeCamp, S. J., Naganathan, A. N., Waldauer, S. A., Bakajin, O. & Lapidus, L. J. (2009). Direct observation of downhill folding of lambda-repressor in a microfluidic mixer. *Biophys. J.* **97**, 1772–1777.
35. Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. & Baker, D. (1999). Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat. Struct. Biol.* **6**, 554–556.
36. Jacob, J., Krantz, B., Dothager, R. S., Thiyagarajan, P. & Sosnick, T. R. (2004). Early collapse is not an obligate step in protein folding. *J. Mol. Biol.* **338**, 369–382.
37. Jacob, J., Dothager, R. S., Thiyagarajan, P. & Sosnick, T. R. (2007). Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J. Mol. Biol.* **367**, 609–615.
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. Welker, E., Maki, K., Shastry, M. C., 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.
40. Kimura, T., Akiyama, S., Uzawa, T., Ishimori, K., Morishima, I., Fujisawa, T. & Takahashi, S. (2005). Specifically collapsed intermediate in the early stage of the folding of ribonuclease A. *J. Mol. Biol.* **350**, 349–362.
41. Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T. *et al.* (2005). Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. *Proc. Natl Acad. Sci. USA*, **102**, 2748–2753.
42. Lapidus, L. J., Yao, S., McGarrity, K. S., Hertzog, D. E., Tubman, E. & Bakajin, O. (2007). Protein hydrophobic collapse and early folding steps observed in a microfluidic mixer. *Biophys. J.* **93**, 218–224.
43. Qiu, L., Zachariah, C. & Hagen, S. J. (2003). Fast chain contraction during protein folding: "foldability" and collapse dynamics. *Phys. Rev. Lett.* **90**, 168103.
44. Mok, K. H., Kuhn, L. T., Goez, M., Day, I. J., Lin, J. C., Andersen, N. H. & Hore, P. J. (2007). A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature*, **447**, 106–109.
45. Pradeep, L. & Udgaonkar, J. B. (2004). Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Biol. Chem.* **279**, 40303–40313.
46. Hofmann, H., Golbik, R. P., Ott, M., Hubner, C. G. & Ulbrich-Hofmann, R. (2008). Coulomb forces control the density of the collapsed unfolded state of barstar. *J. Mol. Biol.* **376**, 597–605.
47. Booker, G. W., Gout, I., Downing, A. K., Driscoll, P. C., Boyd, J., Waterfield, M. D. & Campbell, I. D. (1993). Solution structure and ligand-binding site of the SH3 domain of the p85α subunit of phosphatidylinositol 3-kinase. *Cell*, **73**, 813–822.
48. Guijarro, J. I., Morton, C. J., Plaxco, K. W., Campbell, I. D. & Dobson, C. M. (1998). Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. *J. Mol. Biol.* **276**, 657–667.
49. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
50. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature*, **430**, 586–590.
51. Neudecker, P., Zarrine-Afsar, A., Choy, W. Y., Muhandiram, D. R., Davidson, A. R. & Kay, L. E. (2006). Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J. Mol. Biol.* **363**, 958–976.
52. Li, J., Matsumura, Y., Shinjo, M., Kojima, M. & Kihara, H. (2007). A stable α-helix-rich intermediate is formed by a single mutation of the β-sheet protein, src SH3, at pH 3. *J. Mol. Biol.* **372**, 747–755.
53. Li, J., Shinjo, M., Matsumura, Y., Morita, M., Baker, D., Ikeguchi, M. & Kihara, H. (2007). An alpha-helical burst in the src SH3 folding pathway. *Biochemistry*, **46**, 5072–5082.
54. Wani, A. H. & Udgaonkar, J. B. (2009). Revealing a concealed intermediate that forms after the rate-limiting step of refolding of the SH3 domain of PI3 kinase. *J. Mol. Biol.* **387**, 348–362.
55. Wani, A. H. & Udgaonkar, J. B. (2009). Native state dynamics drive the unfolding of the SH3 domain of PI3 kinase at high denaturant concentration. *Proc. Natl Acad. Sci. USA*, **106**, 20711–20716.
56. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl Acad. Sci. USA*, **95**, 4224–4228.
57. Stryer, L. (1965). The interaction of naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol.* **13**, 482–495.
58. Lillo, M. P. & Beechem, J. M. (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.
59. Haas, E. (2005). The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements. *ChemPhysChem*, **6**, 858–870.
60. Sridevi, K. & Udgaonkar, J. B. (2002). Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry*, **41**, 1568–1578.
61. Ramachandran, S. & Udgaonkar, J. B. (1996). Stabilization of barstar by chemical modification of the buried cysteines. *Biochemistry*, **35**, 8776–8785.
62. Goldenberg, D. P. (2003). Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **326**, 1615–1633.
63. Williams, C., Brochard, F. & Frisch, H. L. (1981). Polymer collapse. *Annu. Rev. Phys. Chem.* **32**, 433–451.
64. Sherman, E. & Haran, G. (2006). Coil-globule transition in the denatured state of a small protein. *Proc. Natl Acad. Sci. USA*, **103**, 11539–11543.

65. Kazmirski, S. L., Wong, K. B., Freund, S. M. V., Tan, Y. J., Fersht, A. R. & Daggett, V. (2001). Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl Acad. Sci. USA*, **98**, 4349–4354.
66. Garcia, P., Serrano, L., Durand, D., Rico, M. & Bruix, M. (2001). NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: Implications for protein folding initiation. *Protein Sci.* **10**, 1100–1112.
67. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T. *et al.* (2002). Long-range interactions within a nonnative protein. *Science*, **295**, 1719–1722.
68. Northey, J. G., Di Nardo, A. A. & Davidson, A. R. (2002). Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol.* **9**, 126–130.
69. Viguera, A. R., Jimenez, M. A., Rico, M. & Serrano, L. (1996). Conformational analysis of peptides corresponding to β-hairpins and a β-sheet that represent the entire sequence of the α-spectrin SH3 domain. *J. Mol. Biol.* **255**, 507–521.
70. Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V. & Eaton, W. A. (2007). Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl Acad. Sci. USA*, **104**, 1528–1533.
71. Ziv, G., Thirumalai, D. & Haran, G. (2009). Collapse transition in proteins. *Phys. Chem. Chem. Phys.* **11**, 83–93.
72. Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A. & Austin, R. H. (1999). Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA*, **96**, 10115–10117.
73. Schonbrunner, N., Koller, K. P. & Kiefhaber, T. (1997). Folding of the disulfide-bonded β-sheet protein tendamistat: rapid two-state folding without hydrophobic collapse. *J. Mol. Biol.* **268**, 526–538.
74. Shastry, M. C. & Udgaonkar, J. B. (1995). The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* **247**, 1013–1027.
75. Patra, A. K. & Udgaonkar, J. B. (2007). Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways. *Biochemistry*, **46**, 11727–11743.
76. Hamadani, K. M. & Weiss, S. (2008). Nonequilibrium single molecule protein folding in a coaxial mixer. *Biophys. J.* **95**, 352–365.
77. Qi, P. X., Sosnick, T. R. & Englander, S. W. (1998). The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nat. Struct. Biol.* **5**, 882–884.
78. Sadqi, M., Lapidus, L. J. & Munoz, V. (2003). How fast is protein hydrophobic collapse. *Proc. Natl Acad. Sci. USA*, **100**, 12117–12122.
79. Pace, C. N. (2001). Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry*, **40**, 310–313.
80. Dyson, H. J., Wright, P. E. & Scheraga, H. A. (2006). The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. USA*, **103**, 13057–13061.
81. Matheson, R. R. & Scheraga, H. A. (1978). Method for predicting nucleation sites for protein folding based on hydrophobic contacts. *Macromolecules*, **11**, 819–829.
82. Parker, M. J. & Marqusee, S. (1999). The cooperativity of burst phase reactions explored. *J. Mol. Biol.* **293**, 1195–1210.
83. Dill, K. A. & Bromberg, S. (2003). Polymer elasticity and collapse. In *Molecular Driving Forces*, chapt. 32, pp. 621–626, Garland Science, New York.
84. Guo, L., Chowdhury, P., Glasscock, J. M. & Gai, F. (2008). Denaturant-induced expansion and compaction of a multi-domain protein: IgG. *J. Mol. Biol.* **384**, 1029–1036.
85. Oliva, F. Y. & Munoz, V. (2004). A simple thermodynamic test to discriminate between two-state and downhill folding. *J. Am. Chem. Soc.* **126**, 8596–8597.
86. Hamada, D., Hoshino, M., Kataoka, M., Fink, A. L. & Goto, Y. (1993). Intermediate conformational states of apocytochrome-c. *Biochemistry*, **32**, 10351–10358.
87. Magg, C., Kubelka, J., Holtermann, G., Haas, E. & Schmid, F. X. (2006). Specificity of the initial collapse in the folding of the cold shock protein. *J. Mol. Biol.* **360**, 1067–1080.
88. Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1995). Is burst hydrophobic collapse necessary for protein folding? *Biochemistry*, **34**, 3066–3076.
89. Miranker, A. D. & Dobson, C. M. (1996). Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.* **6**, 31–42.
90. Klimov, D. K. & Thirumalai, D. (1996). Criterion that determines the foldability of proteins. Phys. *Rev. Lett.* **76**, 4070–4073.
91. Oliveira, L. C., Silva, R. T. H., Leite, V. B. P. & Chahine, J. (2006). Frustration and hydrophobicity interplay in protein folding and protein evolution. *J. Chem. Phys.* **125**, 084904-1-7.