doi:10.1006/jmbi.2000.3618 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **297**, 1217–1229

JMB

Similarities Between the Spectrin SH3 Domain Denatured State and its Folding Transition State

Tanja Kortemme[1]*, Mark J. S. Kelly[1,2], Lewis E. Kay[3] Julie Forman-Kay[4,5] and Luis Serrano[1]

[1]*European Molecular Biology Laboratory (EMBL) Meyerhofstrasse 1 D-6917, Heidelberg, Germany*

[2]*Institute for Molecular Pharmacology, Alfred-Kowalke-Str., D-10315, Berlin, Germany*

[3]*Protein Engineering Network Centres of Excellence and Departments of Chemistry Biochemistry and Molecular and Medical Genetics University of Toronto, Toronto Ontario, M5S 1A8, Canada*

[4]*Structural Biology and Biochemistry Programme Hospital for Sick Children Toronto, Ontario, M5G 1X8 Canada*

[5]*Department of Biochemistry University of Toronto, Toronto Ontario, M5S 1A8, Canada*

**Corresponding author*

We have expanded our description of the energy landscape for folding of the SH3 domain of chicken α-spectrin by a detailed structural characterization of its denatured state ensemble (DSE). This DSE is significantly populated under mildly acidic conditions in equilibrium with the folded state. Evidence from heteronuclear nuclear magnetic resonance (NMR) experiments on ^{2}H, ^{15}N-labeled protein suggests the presence of conformers whose residual structure bears some resemblence to the structure of the folding transition state of this protein. NMR analysis in a mutant with an engineered, non-native α-helical tendency shows a significant amount of local non-native structure in the mutant, while the overall characteristics of the DSE are unchanged. Comparison with recent theoretical predictions of SH3 domain folding reactions reveals an interesting correlation with the predicted early events. Based on these results and recent data from other systems, we propose that the DSE of a protein will resemble the intermediate or transition state of its nearest rate-limiting step, as a consequence of simple energetic and kinetic principles.

© 2000 Academic Press

Keywords: denatured state ensemble; transition state; protein folding; NMR; deuteration

Introduction

Our understanding of how small, single domain proteins fold to their native three-dimensional structures has significantly advanced in the last few years. The protein engineering method (Fersht *et al.*, 1992), and recent approaches using solvent perturbation (Chiti *et al.*, 1999a), have revealed structural properties of the folding transition state ensemble of a growing number of proteins (Jackson, 1998). A major step has been the combination of experimental data with theoretical descriptions of protein folding proceeding along energy landscapes and through folding funnels (Bryngelson *et al.*, 1995; Dobson & Karplus, 1999). Recently, simple models based on native state topology as the major determinant of protein folding mechanisms have successfully captured the essential features of the transition state ensembles observed experimentally (Alm & Baker, 1999a,b; Galzitskaya & Finkelstein, 1999). The unfolded state ensemble of a protein is as important as the folded and transition states in terms of a complete quantitative description of protein folding and stability (Shortle, 1996), but is much less understood. In this regard, simple energetic principles may be helpful in understanding the conformational properties of protein unfolded states.

Conceptually, we must distinguish between the unfolded state of a protein under high denaturant

Present address: T. Kortemme, Department of Biochemistry, University of Washinton, Box 357350, Seattle, WA 98195-7350, USA.

Abbreviations used: DSE, denatured state ensemble; USE, unfolded state ensemble.

E-mail address of the corresponding author: kortemme@u.washington.edu

0022-2836/00/051217–13 $35.00/0

© 2000 Academic Press

conditions (unfolded state ensemble; USE), and the state present under near-native or native conditions (denatured state ensemble; DSE). These so-called ''state'' ensembles can be separated by an energy barrier, in which case they conform to the definition of true thermodynamic states, or they can merely represent a change in the conformational properties of the ensemble as the environmental conditions change (Qi *et al*., 1998). It is the DSE present under non-denaturing conditions that is most relevant to an understanding of the initial steps in the folding process of proteins.

The development of nuclear magnetic resonance techniques with improved spectral resolution (Wüthrich, 1994; Zhang *et al*., 1997a) has led to significant progress in the structural characterization of DSEs (Alexandrescu *et al*., 1994; Gillespie & Shortle, 1997a,b; Mok *et al*., 1999; Sinclair & Shortle, 1999; Zhang & Forman-Kay, 1995, 1997; Zhang *et al*., 1997b). However, while a random-coil model for USEs has been reported (Hennig *et al*., 1999; Smith *et al*., 1996), and approaches for the description of partially folded states are being developed (Shoemaker & Wolynes, 1999), a theoretical model of DSEs in the context of the energy landscape for protein folding has been lacking.

Therefore, detailed experimental information on DSEs is particularly valuable in the context of a well-characterized protein system, into which the properties of DSEs can be integrated with knowledge about the characteristics of the transition state ensemble, the kinetics of folding, and the energetic properties of the native state. A well-suited model protein for this purpose is the SH3 domain of chicken α-spectrin. Its native structure (Blanco *et al*., 1997; Musacchio *et al*., 1992), its two-state folding mechanism and transition state ensemble (Viguera *et al*., 1994), the role of topology in the folding process (Viguera *et al* 1996b), and the effect of changing local secondary structure propensities on the kinetics and thermodynamics (Prieto *et al*., 1997) have all been experimentally characterized. Moreover, the very similar global features of the transition state ensemble observed for the SH3 domain from *src* (Riddle *et al*., 1999) suggest that rate-limiting steps in folding could be largely dependent on topology (Alm & Baker, 1999a), and could therefore be common to a number of proteins sharing the same fold (Chiti *et al*., 1999b; Martinez & Serrano, 1999; Riddle *et al*., 1999).

The folded state of the chicken α-spectrin SH3 domain has an orthogonal β-sheet sandwich fold, with five unstructured residues at the N terminus. The folded state and DSE of this domain are in slow exchange on the NMR time-scale, and can be observed simultaneously under mildly acidic conditions. Heteronuclear NMR methods, with improved spectral resolution tailored to the study of protein DSEs based on the observation of NOEs up to an apparent average distance of about 5 Å, indicated that the spectrin SH3 domain DSE had little residual structure (Blanco *et al*., 1998). However, an important question remained unsolved: is the lack of observed long-range information a consequence of the largely extended conformation of the investigated DSE or is it due to experimental limitations? Evidence for the relative compactness of some DSEs was inferred from small-angle X-ray scattering (Kataoka & Goto, 1996) and paramagnetic relaxation studies, yielding the first detailed structural description of a DSE (Gillespie & Shortle, 1997a,b). Protein deuteration has been shown to allow the detection of NOEs between labile protons over distances significantly longer than 5 Å, as a consequence of reduced relaxation (Mal *et al*., 1998; Mok *et al*., 1999; Pachter *et al*., 1992). Here, this strategy is employed, with the aim of obtaining a previously unattained level of detail in the description of the DSE under near-native conditions for the well-characterized α-spectrin SH3 domain. Moreover, we are interested in how the conformational properties of the DSE shift by introducing strong non-native local interactions. For this reason, we have compared our results on the wild-type (WT) spectrin SH3 domain to those obtained for a triple mutant (AE-K: Thr4 to Ala, Gly5 to Glu and Aspl4 to Lys) previously designed to study the influence of non-native α-helical tendency engineered into the N terminus on protein folding. The folded state structures of the AE-K and WT protein are identical, although the AE-K protein is less stable (Prieto *et al*., 1997). We observe native-like residual structure in the DSE of both proteins, whose location broadly coincides with regions found structured in the transition state for folding of the spectrin SH3 domain. This offers intriguing insight into the energy landscape for the folding of this small β-sheet domain.

Results

Protein perdeuteration strategy

The conformational properties of the DSEs of the ^{2}H, ^{15}N-labeled WT spectrin SH3 domain and the AE-K protein under mildly acidic conditions were studied by NMR ^{1}H-^{15}N HSQC-NOESY-HSQC experiments (Frenkiel *et al*., 1990; Ikura *et al*., 1990). The greater level of dispersion of the ^{15}N signals facilitates the assignment of the resonances in the DSE (Zhang *et al*., 1997a), while deuteration enables much more efficient NOE transfer, allowing NOEs to be detected between protons at greater apparent average distance than the 5 Å typically observed (Mok *et al*., 1999). In unfolded states, we would expect that a greater efficiency of NOE transfer would enable more short distance correlations to be observed, such that the averaging and dynamics that reduce NOE intensity would not completely destroy the NOE. This approach has three limitations: (i) perdeuteration restricts the number of protons that can be observed to only those which exchange back to protons in water. In our case, only the backbone amide groups and the indole protons of the two tryptophan residues were used in the NOE analysis, as other exchangeable side-chain

groups were not assigned in the DSE. (ii) During the long mixing times (up to 600 ms) used to obtain long-range information, there is time for the protein to fold and unfold, with an approximate rate of 12 s^{-1} for folding and 0.08 s^{-1} for unfolding at pH 3.5 and 298 K, in the case of the WT protein (Martinez *et al.*, 1998) (similar numbers are obtained at low temperatures, data not shown). As a consequence, NOEs observed between two proton resonances in the DSE (NOE^D, see Scheme 1) can originate from the native conformation (NOE^N) as a result of folding and unfolding processes (EX) during the mixing time (Scheme 1).

$$\begin{array}{ccc} A^N & \xleftrightarrow{NOE^N_{A-B}} & B^N \\ \updownarrow EX^A & & \updownarrow EX^B \\ A^D & \xleftrightarrow{NOE^D_{A-B}} & B^D \end{array}$$

Scheme 1.

Native-like NOEs in the DSE (observation of an $A^D - B^D$ NOE, when the $A^N - B^N$ NOE is also present) are taken to arise from the DSE if the corresponding exchange NOEs (NOE + EX) between protons A and B in native and denatured states, respectively, and *vice versa* ($A^D - B^N$ and $A^N - B^D$), are not observed (Mok *et al.*, 1999). Our rationale is that a three-step process $EX^A - NOE^N - EX^B$, which would lead to an NOE observed between resonance of the DSE, but originating from the native state, is extremely unlikely if the corresponding two-step processes $EX^A - NOE^N$ or $NOE^N - EX^B$, which give rise to exchange NOEs of type $A^D - B^N$ or $A^N - B^D$, are not observed. Here, most native-like NOEs observed in the DSE are considered to be representative of interactions within the DSE, as the simultaneous presence of unambiguous native-like NOEs in the DSE and of the corresponding exchange NOEs ($A^D - B^D$, $A^N - B^N$, $A^D - B^N$, and $A^N - B^D$) only occurs for three NOEs involving the distal hairpin (marked by asterisks in Figure 4). (iii) The detection of multiple NOEs does not allow discrimination between the presence of one major conformation characterized by all the NOE-derived short distances, or an ensemble of many different conformers which could have a single or a small subset of short distances in each one. As the second scenario is more likely, we emphasize that our data describe an ensemble of conformations, as expected for the denatured ''state''.

Figure 1 shows the number of unambiguous NOEs observed in the folded states of the spectrin WT and AE-K SH3 domains classified by interproton distance in the crystal structures, demonstrating that the perdeuteration strategy allows the observation of long-range NOEs, up to approximately 10 Å. The efficient transfer of magnetization

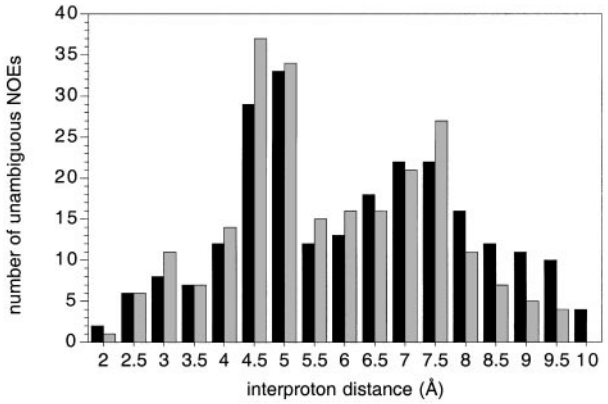

Figure 1. Histogram of unambiguous NOE crosspeaks observed at various interproton distances in the folded state of the ^2H, ^{15}N-labeled spectrin SH3 domain WT (black bars) and AE-K mutant (gray bars) using a mixing time of 600 ms at 278 K. Interproton distances are grouped together in bin x for distances of $x - 0.5$ to x Å on the basis of interproton distances in the folded structure.

to protons close in space *via* NOE rather than loss *via* spin diffusion or relaxation is a prerequisite for observation of long-range interactions in the DSE. It also allows comparison to the unfolded state of the N-terminal SH3 domain of the *Drosophila* drk protein (drkN SH3 domain U_{exch} state), where NOEs reflecting similar distances were detected in the folded state (Mok *et al.*, 1999). While the influence of spin-diffusion on signals in the spectra at the mixing times used cannot be excluded, the contribution of these effects have been significantly reduced through the use of high levels of deuteration (Mal *et al.*, 1998; Mok *et al.*, 1999).

Medium and long-range NOEs in the DSEs of WT and AE-K proteins

Representative strips of the 600 ms mixing time ^1H-^{15}N HSQC-NOESY-HSQC spectra for the WT and AE-K proteins displaying medium and long-range NOEs in the DSE are shown in Figure 2. In the AE-K mutant, NH-NH NOEs between residue 3 and residues 6 or 7 in the DSE indicate the presence of the engineered α-helical conformation in the N-terminal region, which is absent in the WT protein (Figure 2(a)). Medium and long-range NOEs found for residues 9, 24, 43 and 53 in the WT DSE are displayed in Figure 2(b). Summaries of the observed NOEs in the DSE are given in Figure 3(a) for AE-K and (b) for WT.

In the WT protein DSE, we have observed several native and non-native medium-range NH-NH NOEs involving loop and turn regions around regions 14 to 23 and 24 to 34, as well as from the side-chain indole group protons of the tryptophan residues 41 and 42 (Figure 4(a)). Moreover, several

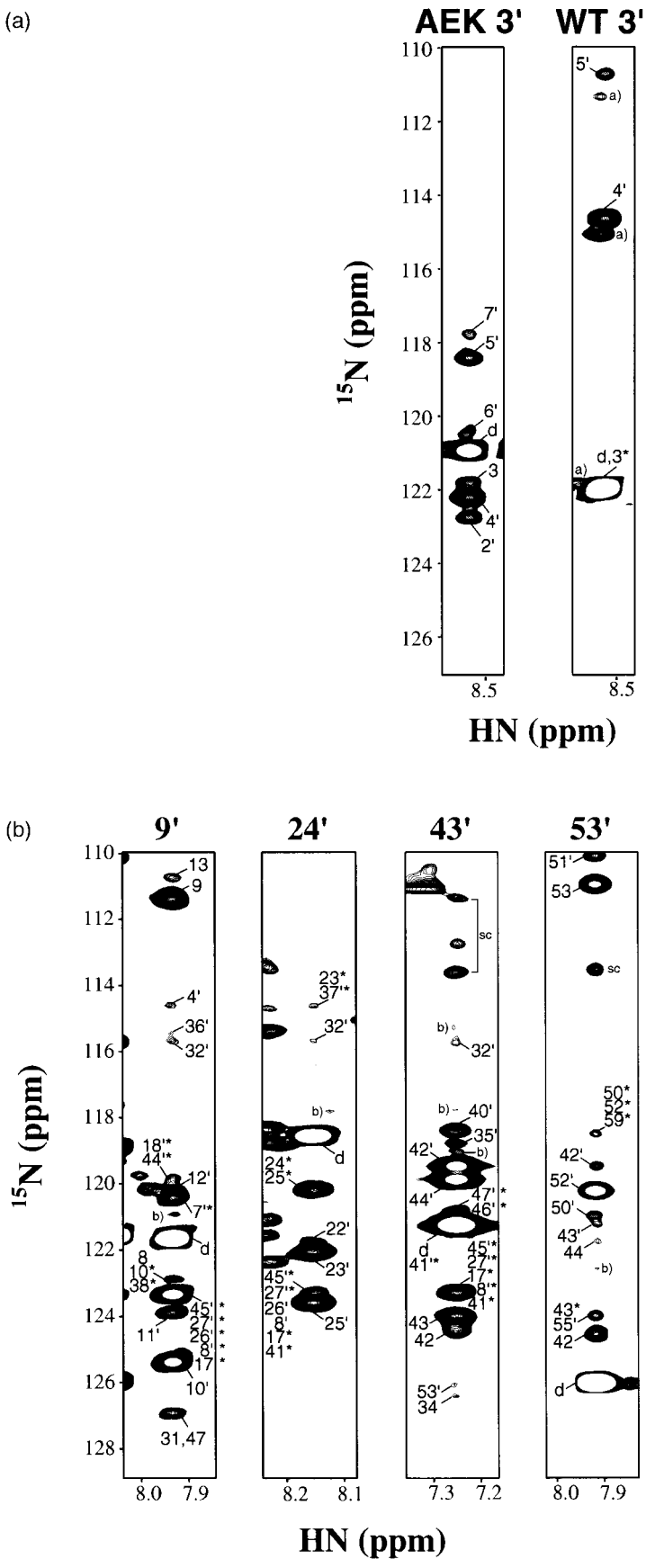

Figure 2. (*Legend opposite*)

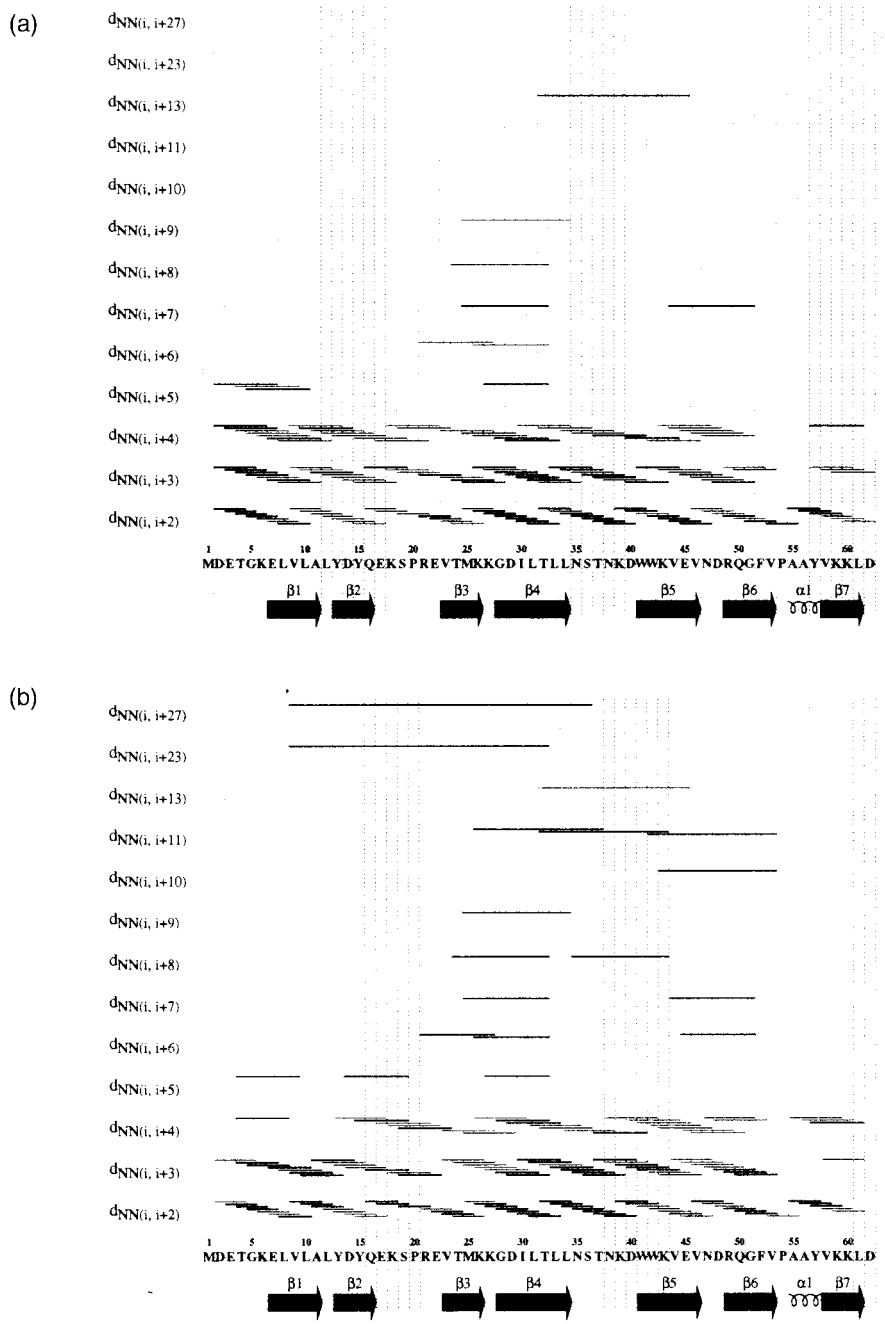

Figure 3. Summary of NH-NH NOEs observed in the DSEs of the (a) AE-K mutant and (b) WT. Line width does not relate to the intensity of the NOEs. Gray lines represent possible NOEs that cannot be unambiguously assigned or ruled out due to signal overlap. For NOEs $\leqslant (i, i+4)$ all possible NOEs were examined; absence of a line for these NOEs indicates the lack of intensity for at least one of the two symmetry-related positions.

Figure 2. Strips of a ^{1}H-^{15}N HSQC-NOESY-HSQC spectrum (mixing time 600 ms, 278 K) taken at the ^{15}N and ^{1}H chemical shifts of the indicated residues in the DSE of the spectrin SH3 domain (a) AE-K mutant and WT, and (b) WT. An sc denotes NOEs to side-chain NH_2 groups of Asn and Gln residues. A prime following the residue number, denotes an NOE from the DSE, while a residue number lacking this indicates an NOE from the folded state. An asterisk indicates that an NOE could not be unambiguously assigned due to overlap at both symmetry-related resonance positions. A d indicates the diagonal peak. An a) marks crosspeaks originating from an unrelated resonance and b) marks signals lacking symmetry-related intensity.

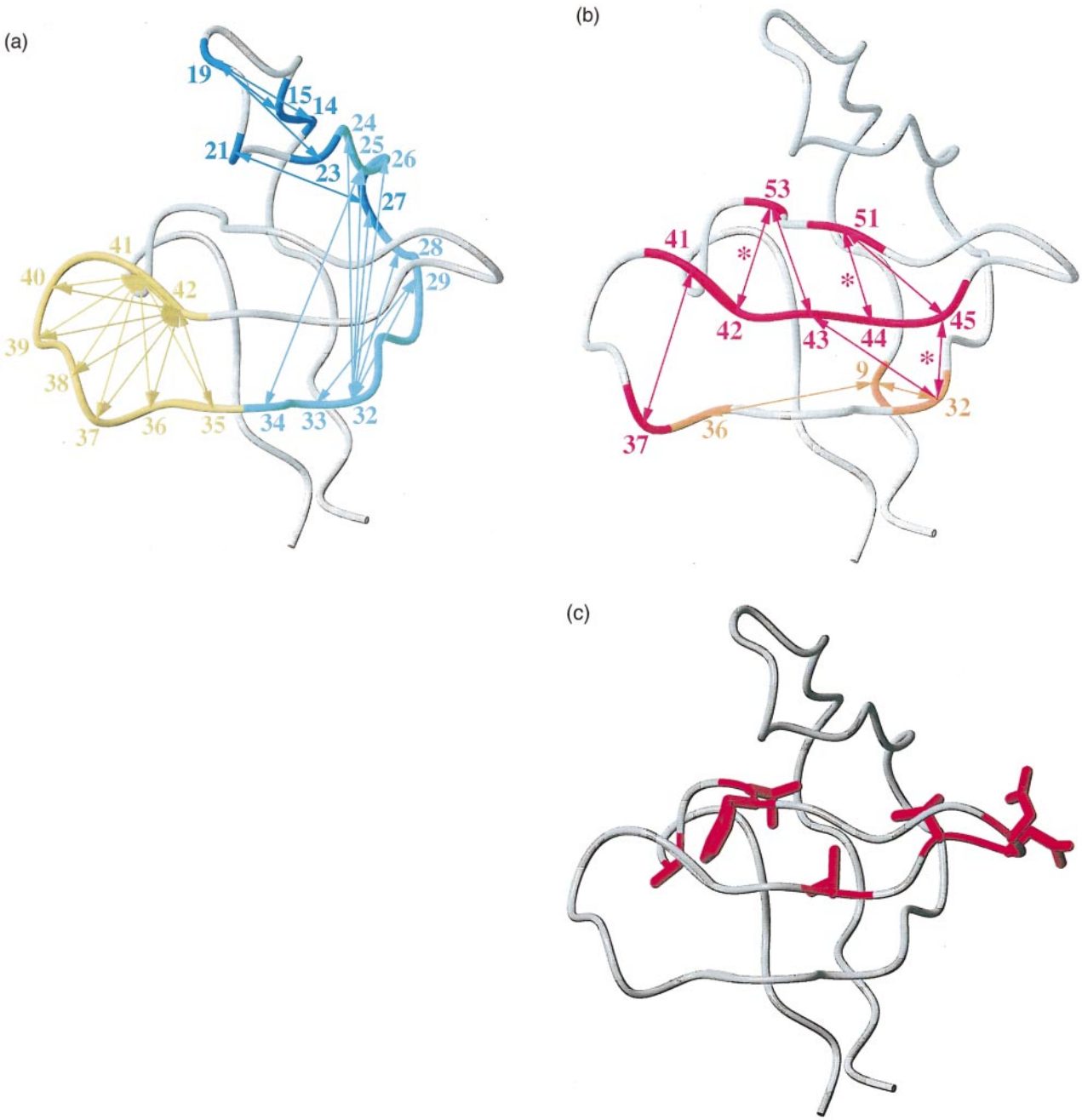

Figure 4. NH-NH NOEs observed in the DSE of the WT spectrin SH3 domain, depicted on a ribbon representation of the structure of the folded state. (a) Primarily local NH-NH NOEs (blue and cyan) and NH-tryptophan indole NOEs (yellow, involving W41 and W42). (b) Medium and long-range NH-NH NOEs in the region of the central three-stranded β-sheet in the folded structure (red) and connecting the N-terminal region to one of the outer strands of the central sheet (orange). (c) Residues used for the ''protein engineering'' analysis that were found to be part of the folding nucleus formed in the transition state in the spectrin SH3 domain (red) (Martinez & Serrano, 1999). In that study there were no mutant probes between residues 36 and 43, and therefore no information is available about the transition state conformation of this region which was found to have interactions in the DSE.

native-like medium and long-range NOEs are found within the central β-sheet region, which are especially prominent in the distal β-hairpin (residues 41 to 53) (Figure 4(b)). All of these long and medium-range NOEs have low intensities (Figure 2(b)), indicating either long distances between the protons involved or, more probably, a low population of the conformers with short distances between protons, giving rise to the respective NOE.

While NOEs indicative of the non-native engineered helical region were observed in the DSE of

AE-K, we could not detect evidence for interactions between this helix and the rest of the protein. The possibility of such interactions was suggested by the discrepancy in the helical population estimated from $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ chemical shifts in the DSE of AE-K (~50 %; Blanco *et al.*, 1998) and in an isolated peptide comprising the helical region (~24 %; Prieto *et al.*, 1997), and was supported by chemical shift perturbations of resonances of residues 30 to 36 in the AE-K mutant compared to the WT (Blanco *et al.*, 1998). However, NOEs between residues 9 and 32 (native-like NOE, defined as distance <10 Å in the folded protein) and 9 and 36 (non-native, distance >10 Å in the folded protein) were detected in the WT DSE (Figures 2(b) and 3(b)). This confirms the previous conclusion of contacts between these regions in the DSE, although we cannot discriminate between whether the interaction with the non-native helical region involving residues 2 to 10, or disruption of these contacts by a non-native helix, causes the change in the chemical shifts around residues 30 to 36 previously found.

Apart from the differences in the N-terminal region, which are to be expected from the non-native helical propensity, there is no evidence for significant changes in the DSE of the AE-K mutant compared to that of WT. The larger number of NOEs detected in the WT DSE is most easily explained by the higher sample concentration (2.6 mM DSE for WT *versus* 1.3 mM DSE for AE-K) due to the poorer solubility of the latter protein. The low intensity of the 9-32 and 9-36 NOEs of the WT shown in Figure 2(b) makes it possible that similar contacts are present within the AE-K DSE, but cannot be observed due to the lower sample concentration. The situation is similar for the folded states of both proteins, where we can detect a significantly larger number of long-range NOEs above 8 Å distance in the case of WT (Figure 1), although both proteins have identical native structures (Prieto *et al.*, 1997).

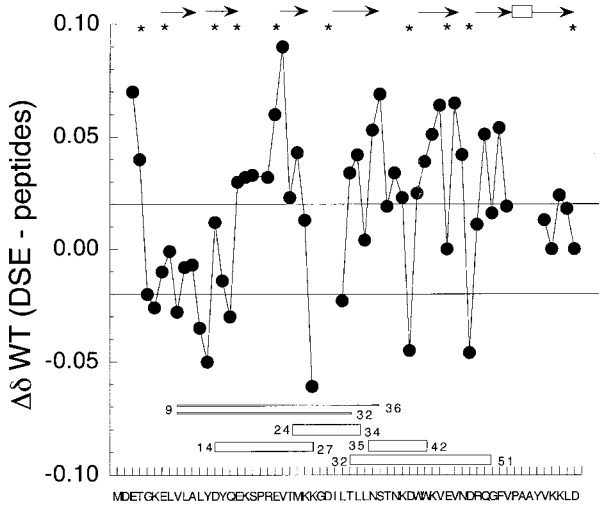

Figure 5. Conformational shift difference between the $^{1}H^{\alpha}$ protons of the WT DSE and a series of peptides encompassing the entire sequence of the SH3 domain (Viguera *et al.*, 1996a). The asterisks indicate the Asp and Glu residues which could exhibit different values due to the fact that the pH used for the peptide analysis is not exactly identical in all cases with that used to study the DSE of the WT protein. Horizontal lines mark the range of very small chemical shift deviations (±0.02 ppm). The boxes at the bottom of the Figure indicate the regions where long-range NOEs have been detected. The bars show the two long-range NOEs involving non-native interactions. Secondary structure is shown at the top of the Figure. Arrows correspond to β-strands and the box to the 3_{10}-helix.

Low population of structured conformers in the DSEs of WT and AE-K proteins

Previous comparison of the chemical shifts of the $^{1}H^{\alpha}$ protons of the DSE of the spectrin WT SH3 domain, obtained from a protonated sample, to those of a peptide corresponding to the first 19 residues revealed no significant differences (Blanco *et al.*, 1998). Figure 5 shows a comparison of the chemical shifts of the whole domain to those of a series of peptides (Viguera *et al.*, 1996a). The majority of the values differ by less than 0.05 ppm, again indicating that the structured compact conformers in the DSE of the spectrin SH3 make up a relatively small fraction of the ensemble. However, the measured small differences of the chemical shift values for the central three-stranded β-sheet region (residues 32 to 53) to the isolated peptides are as expected if a small population of the molecules in the protein DSE has a preference for native-like secondary structure (positive chemical shift deviations for the β-strands and negative for the β-turns). In contrast, the chemical shift perturbations of the N-terminal strand in the WT DSE *versus* the isolated peptides indicate non-native (turn-like or helical) tendencies, in agreement with the observation of the non-native (i, $i+4$) and (i, $i+5$) NOEs involving residue 4, and the non-native NOE between residues 9 and 36. Thus, the regions for which we find several medium and long-range NOEs coincide with those having small, but significant differences between the H^{α} proton chemical shifts of the peptides and those of the protein.

Discussion

Similarities between the DSE of the spectrin SH3 domain and its folding transition state

The DSE of the spectrin SH3 domain appears from initial inspection of the data to be quite unstructured. Local conformations are extensively averaged, as suggested from the simultaneous presence (or overlap) of nearly all sequential and short-range NH-NH NOEs (i, $i+4$) (Figure 3). Of the limited number of medium and long-range interactions observed, many NOEs (i, $i+7$) are

centered around the distal β-hairpin (residues 41 to 53) and the preceding strand (residues 32 to 37) (Figure 4(b)). This suggests a preferential sampling of native-like topology in the central three-stranded β-sheet of the spectrin SH3 domain, while the N and C-terminal regions (residues 1 to 8 and 54 to 62) do not display any non-local NOEs. This is reminiscent of the distribution of structure in the folding transition state of the spectrin SH3 domain, as inferred using the protein engineering method (Martinez & Serrano, 1999), where data suggest an almost complete formation of the distal β-hairpin (residues 41 to 53) in the folding transition state. Figure 4(c) depicts the side-chains of residues involved in significant interactions in the folding transition state including turn-like interactions around residues 47 and 48 and native-like interactions of hairpin residues 44, 46, 52 and 53. The formation of the distal β-hairpin appears to be an obligatory step in the folding of the spectrin as well as the *src* SH3 domain, which shares the same topology but has only 36% sequence identity (Riddle *et al.*, 1999). While the transition state analysis using the protein engineering method (Fersht *et al.*, 1992) reports on the presence of interactions involving side-chains in the kinetic transition state, the NMR-based strategy applied here yields information on the protein backbone in the equilibrium DSE. Nevertheless, both methods are indicative of the presence of a hairpin-like structure. All cross-strand NH-NH NOEs expected for distal β-hairpin formation (closer than 5 Å in the native structure) are observed unambiguously (42-53, 43-53, 44-51, 45-51) or overlapped (44-52, 46-50 and 46-51) in the DSE, although the NMR data do not allow any conclusion about the simultaneous presence of all interactions (see below).

The apparent similarity of the kinetic transition state and equilibrium DSE is additionally supported by observations for other regions of the protein. The transition state of the spectrin SH3 domain appears to have non-native interactions involving the side-chain of residue 33, and we observe a number of non-native NOEs in the region 24 to 34 in the spectrin SH3 domain DSE (Figure 4(a)). While there is no experimental information about the transition state conformation in the region 36 to 43 in the spectrin SH3 domain, comparison with the *src* SH3 domain also suggests similarities, assuming that the *src* and spectrin SH3 domain transition states resemble each other, as posited by Martinez & Serrano (1999) and Riddle *et al.* (1999). Data for *src* implicate a portion of the strand N-terminal to the nucleating β-hairpin as part of the folding transition state (Riddle *et al.*, 1999). This is in good agreement with the observed NOEs in the spectrin SH3 domain DSE between this strand and the distal hairpin (NOEs 32-43 and 37-41, see Figure 4(b)).

The finding that structural features of the DSE of a protein could resemble its folding transition state raises a potential conceptual problem. One of the basic assumptions of the protein engineering analysis of transition state structure is that the energetics of the DSE are unchanged by mutation. Our data, however, suggest similar interactions stabilizing both the DSE and transition state, which should therefore both be affected by mutations located in the structured regions. However, the population of structured species in the spectrin DSE is very small. Therefore, in practical terms the DSE can be viewed in this protein as an ensemble of rapidly interconverting conformers with small differences in energy. Thus, a mutation destabilizing a native-like interaction in the DSE will shift the properties of the DSE by decreasing the population of those conformers containing the interaction, but it will not cause significant energetic changes affecting transition state analysis. However, in cases such as the drk SH3 domain, where a single dominant compact conformation may be present in the DSE, mutations could have significant effects on DSE structure and stability (Mok *et al.*, 1999 and unpublished results).

Single dominant or multiple conformations in the spectrin SH3 domain DSE

Although the NMR techniques applied here allow the observation of long-range NOEs, even the transition-state-like features detected in the DSE are mostly described by medium-range NOEs between adjacent β-strands or across loops. Additionally, the experimental method does not allow us to conclude whether or not all or a large number of non-local NOEs originate from a single, dominant conformation. In fact, the more plausible scenario is the simultaneous presence of many conformations, with some displaying fluctuating characteristics resembling parts of the folding nucleus. A similar conclusion of many coupled local interactions being present in a DSE in the almost complete absence of long-range interactions was recently inferred from backbone dynamics measurements in a deletion mutant of staphylococcal nuclease (Sinclair & Shortle, 1999). Together, these experimental data on the DSEs of proteins suggest that the development of a theoretical description of the DSEs for many proteins is feasible. For this, the stabilities of relatively few medium-range interactions with low populations would be taken into account, with the majority of conformers dominated by local interactions.

Of note, molecular dynamics simulations have been applied to describe the thermally denatured state of barnase (Bond *et al.*, 1997). While the behavior seen in the simulations was in agreement with the observations from NMR studies, several predicted tertiary interactions could not be observed in these experiments which were not tailored to detect long distances (Bond *et al.*, 1997). Experimental data such as these resulting from our studies on the DSE of the spectrin SH3 domain might be a good test case for the validity of such molecular dynamics simulations. In fact, molecular dynamics simulations of the *src* SH3 domain

unfolding showed persistent residual structure within the distal hairpin and contacts to the preceding strand (Tsai *et al.*, 1999), similar to the regions where we observe medium and long-range NOEs in the spectrin SH3 domain DSE.

Comparison with drkN SH3 domain

A recent extensive structural characterization of the DSE of a ^{2}H, ^{15}N-labeled sample of the N-terminal SH3 domain of the drk protein (drkN SH3 domain U_{exch} state) revealed a significantly larger number of long-range NOEs than observed here in the α-spectrin SH3 domain (Mok *et al.*, 1999). This could be due to the drkN SH3 domain U_{exch} state being more compact on average, or because of more favorable relaxation properties. The fact that the number of NOEs reflecting distances of 8-10 Å in the folded state shows no significant differences between the two proteins (Mok *et al.*, 1999) favors the first hypothesis. However, diffusion measurements indicate a similar level of compactness on average for both proteins (Mulder & L.E.K., unpublished data). A possible explanation for the difference is the coexistence of at least two different states, separated by an energy barrier, in the drkN SH3 domain U_{exch} state, one of which is more compact (Mok *et al.*, 1999). In this way, the average diffusion rate will be similar to the spectrin SH3 domain, but the observed number of NOEs would be greater. The drkN SH3 domain is significantly less stable than the spectrin and *src* SH3 domains. Recent kinetic and thermodynamic data on mutants of the drkN SH3 domain indicate that part of this difference in stability could be due to a relatively stable compact denatured state (Mok *et al.*, unpublished results). Comparison of the spectrin and drkN SH3 domains makes clear that related proteins do not necessarily possess similar DSEs; however, this does not prevent them from attaining the same native fold.

Distribution of observed medium and long-range NOEs in the spectrin SH3 domain DSE

A number of the medium and long-range NOEs depicted in Figure 4 arise from protons with resonance positions particularly well resolved in the spectra in either or both of the ^{15}N and ^{1}H dimensions (backbone amide groups of Gly, Ser and Thr and Trp residue side-chain indole groups). This facilitates unambiguous NOE assignments and could lead to a possible bias towards NOEs involving these resonances. However, we observe both medium and long-range NOEs arising from backbone amide protons not belonging to Gly, Ser or Thr residues (i.e. Met25-Lys34, Trp42-Val53 and Lys43-Val53), as well as well-resolved resonances not displaying any long-range NOEs. While preferential observation of NOEs involving resolved resonances and failure to unambiguously assign NOEs due to spectral overlap are somewhat limiting, we believe that the picture emerging from all observed NOEs is valid. In the case of the drkN SH3 domain, similar topologies of the DSE were calculated when all, or a randomly chosen 40 %, 60 % or 80 % of the observed NOEs were included (Mok *et al.*, 1999). Thus, the observation of all, or just a subset of NOEs, does not appear to change our understanding of the overall topological features of the DSE presented here.

Implications for folding

It has been suggested that the initial events in protein folding encompass an uphill conformational search for a transition state nucleus in the DSE requiring at least milliseconds (Sosnick *et al.*, 1996). Once enough tertiary interactions are attained to stabilize the folding nucleus in the transition state, the protein passes over the transition state barrier, and the rest of the folding process proceeds downhill in the absence of intermediates (Creighton 1995; Fersht, 1995). If there is an energy barrier in protein folding with the so-called ''transition-state'' at its highest point, it may be considered advantageous for the protein to have this barrier as high as possible to minimize spontaneous unfolding. On the other hand, the barrier should be as low as possible to facilitate fast folding. As long as the folding nucleus is native-like, any mutation that stabilizes it will also stabilize the native state to the same extent, and the DSE to a lesser extent. An excessive stabilization of the folding nucleus (100 % formation in the absence of the additional tertiary contacts formed in the transition state) will result in the folding nucleus being 100 % formed in the DSE and therefore no increase in folding speed. In the opposite case, a very unstable folding nucleus will require many additional tertiary interactions for its stabilization, paying a large entropic penalty, thus slowing the folding reaction. As a result, for efficient folding it would be conceptually expected that the folding nucleus of a protein should be partly populated in the DSE, but only to a small extent (Fersht, 1995). This situation is observed here experimentally, suggesting that transition state-like conformations are sampled in the DSE of the spectrin SH3 domain.

Based on these considerations, conformations resembling the folding nucleus present at low populations in the DSE of a protein should be generally advantageous for protein folding. A series of protein fragments of increasing length from chymotrypsin inhibitor (CI-2) and barnase have been studied in the absence of denaturant (de Prat Gay *et al.*, 1995; Neira & Fersht, 1999). Although no long-range NOEs could be detected within the fully protonated peptides, the authors could identify the existence of transient secondary structures within the peptides which were also present in the folding intermediate of barnase and the transition state of CI-2 (Neira & Fersht, 1999). These results, although not obtained with the full-length proteins, provide support for the presence of conformations

within the DSE that resemble the folding transition state. It is interesting that this has been observed for both a protein with a delocalized transition state involving several different regions of the sequence (CI-2) and one with a largely polarized transition state primarily concentrated around the distal hairpin (SH3). These results are not incompatible with the fact that non-native conformations can be significantly populated in DSEs, as we report for the AE-K mutant, since these involve regions that are not involved in the rate-limiting step.

The most likely explanation for the observed resemblance of the DSE and transition state ensemble is given by simple energetic considerations. In order to provide a significant favorable enthalpic contribution, the interactions formed in the transition state have to be energetically preferred. Thus, they will be populated in the DSE to an extent dependent upon their favorable interaction energy. In other words, if these interactions were not favorable enough to be present in the DSE (although, for the kinetic reasons described above, at low levels) they could not contribute towards the energy required to pass the transition state barrier.

Predictions for folding and the SH3 domain DSE

Alm & Baker (1999b) predicted the folding pathway of the *src* SH3 domain using a simple method based on native-state structures. The model considers contiguous stretches of residues ordered as in the native structure, with all other residues disordered. The free energy is related to the surface area buried, the number of ordered residues and to the length of the loop between the two stretches. With this simple approximation, they found that early in the folding pathway the RT loop (~residues 16 to 26) and the distal loop (~residues 42 to 52) are the first regions to fold. Later on (near halfway through the process), the RT loop partly melts (keeping only a small region around residues 21 to 30), while the distal loop consolidates from an energetic point of view. Interestingly, these regions roughly correspond to those where we find NH-NH NOEs > $(i, i+4)$ in the spectrin SH3 domain DSE (14-19; 21-27; regions 24 to 34 and 42 to 53). Although the force field used by the authors is quite simple, it nevertheless supports our argument that the most likely explanation for the observed resemblance of the DSE and transition state ensemble emerges from simple energetic considerations.

Conclusions

The structural characterization of the denatured state ensemble of the chicken α-spectrin SH3 domain reveals a similarity between the DSE and the transition state for folding. However, conformers lacking significant native-like long-range interactions appear to predominate in the DSE, consistent with the idea that preformed stable ''folding initiation sites'' are disadvantageous for folding. While, in general, the presence of conformers at equilibrium does not confer proof of their kinetic importance, the correlation observed here is in accordance with simple energetic principles and is indicative of how the amino acid sequence codes for efficient folding of a number of proteins.

Materials and Methods

Expression and purification of uniformly ^{2}H, ^{15}N-labeled spectrin SH3 domain

Uniformly ^{2}H, ^{15}N-labeled chicken α-spectrin SH3 domain (WT) and the T2A, G5E, D14K triple mutant protein (AE-K) were expressed using a minimal medium prepared in 99.98 % $^{2}H_2O$ (Ontario Hydro, Tiverton, Ontario), supplemented with 3 g/l [*U*-^{2}H]-glycerol (Martek Biosciences Corporation, Columbia, MD, USA) and 1 g/l $^{15}NH_4Cl$ (Martek). In detail, the minimal medium contained per liter: 8 g Na_2HPO_4, 2 g KH_2PO_4, 0.5 g NaCl, 120 mg $MgSO_4$, 30 mg $CaCl_2$, 1 mg thiamine, 1 mg biotin, 100 μg carbenicillin, 50 mg EDTA (disodium salt), 5 mg $FeCl_3$, 0.5 mg $ZnCl_2$, 0.1 mg $CuCl_2$, 0.1 mg $CoCl_2.6H_2O$ and 0.1 mg H_3BO_3. The $p^{2}H$ was adjusted to 7.0 (pH electrode reading not corrected for deuterium isotope effects). For maximum deuteration, all components of the medium except $^{15}NH_4Cl$, thiamine, biotin and carbenicillin, were dissolved in $^{2}H_2O$ and lyophilized prior to addition to the medium.

The cDNA encoding the SH3 domain from chicken α-spectrin was cloned in a pET3d expression vector (Novagen). The *Escherichia coli* host strain BL21(DE3) was freshly transformed and plated on LB/$^{2}H_2O$ plates supplemented with carbenicillin (100 μg/ml). A single colony was inoculated into 1 ml of the supplemented minimal medium and grown with shaking at 200 rpm at 37 °C. The culture was successively diluted into larger volumes, such that the apparent absorbance at 600 nm A_{600} never exceeded a value of 0.6, until the final 1 l volume of culture reached A_{600} 0.6-0.8. At this point, protein production was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM, and the cells were harvested after a further six hours of shaking at 37 °C. Protein purification was performed essentially as described previously (Blanco *et al.*, 1998). Final protein yields for WT and AE-K were 50 and 11 mg per liter of culture, respectively. The completeness of the isotopic enrichment was assessed by mass spectrometry, taking into account that all exchangeable deuterons were replaced with protons during the purification process. Isotopic enrichment was greater than 99 % in all cases.

NMR spectroscopy

NMR samples were prepared by dissolving the appropriate amount of lyophilized ^{2}H, ^{15}N-labeled protein in a $H_2O/^{2}H_2O$ 9:1 (v/v) mixture containing 1 μM leupeptin. Sample concentrations were 4.3 mM (WT) and 2.0 mM (AE-K), and the pH was adjusted to 2.2 or 3.5 (uncor-

rected for deuterium isotope effects) for WT or AE-K, respectively.

NMR experiments were performed at 278 K on a Varian Inova 500 MHz spectrometer equipped with a *z*-axis pulsed field gradient and an actively shielded triple-resonance probe. Resonance assignments were facilitated by the previously published chemical shifts at 298 K (Blanco *et al*., 1998). Assignments were essentially the same after corrections to account for the different temperature used here and the effect of the deuterium isotope shifts. To detect HN-HN NOEs, ^{1}H-^{15}N HSQC-NOESY-HSQC experiments (Frenkiel *et al* ., 1990; Ikura *et al*., 1990) were acquired as described by Zhang *et al*. (1997a), with data set matrices of $64 \times 44 \times 512$ complex points and spectral widths of 1000, 1000 and 8000 Hz (*F*1, *F*2, *F*3), using a mixing time of 600 ms. This long mixing time was chosen to allow the observation of weak NOEs arising from structured conformations in the DSE present at low populations. The effects of spin diffusion are minimized through the use of perdeuteration (Mal *et al*., 1998; Mok *et al*., 1999). A total of 16 scans were acquired for each FID. The ^{1}H and ^{15}N carrier frequencies were placed at 4.79 and 119.77 ppm, respectively. The ^{1}H chemical shifts were referenced to external DSS, and ^{15}N was referenced indirectly according to Wishart *et al*. (1995).

Data were processed using the NMRPipe/NMRDraw software (Delaglio *et al*., 1995), and analyzed using the program NMRView (Johnson & Blevins, 1994). For the HSQC-NOESY-HSQC experiments, forward-backward linear prediction (Zhu & Bax, 1992) was used to double the size of the time domain in both ^{15}N dimensions. Either 65° shifted squared sine-bell or adjusted Gaussian window functions were applied to all three dimensions. After zero filling and extraction of the data to retain 10.43 to 6.43 ppm of the acquisition dimension, 3D data sets of $512 \times 256 \times 256$ real points were obtained.

Monomeric state of the proteins

Sedimentation equilibrium experiments were performed for ^{1}H WT (pH 2.2) and ^{1}H AE-K (pH 3.5) samples at concentrations of 1.4 and 0.14 mg/ml in the presence of 5 mM citrate buffer, using a Beckman XL-A ultracentrifuge equipped with an An-50 Ti rotor. Samples were run at speeds of 35,000 and 40,000 rpm until equilibrium was reached. The fitting of the equilibrium radial concentration distribution to an ideal single-component model using a partial specific volume calculated from the amino acid sequence (0.741×10^{-3} m^3 kg^{-1} for WT and 0.728×10^{-3} m^3 kg^{-1} for AE-K), yielded molecular masses confirming the monomeric state of both proteins, which was independent of concentration. Additionally, observed NMR line-widths and diffusion measurements (Mulder & L.E.K., unpublished data) are consistent with both proteins being monomeric under the NMR conditions.

Acknowledgments

We thank Ranjith Muhandiram for help with NMR experiments, Catherine Zwahlen and Geoff Mueller for NMRView scripts, and David Baker, Francisco Blanco, Henry Mok, Gerlind Wallon and members of the Kay, Forman-Kay and Serrano laboratories for discussion. This work was supported by funds from the European Union (grant EU CT96-0013) to L.S. and from the Medical Research Council of Canada to J.D.F.-K. T.K. acknowledges the Boehringer Ingelheim Fonds for a Short-Term Travel Fellowship.

References

Alexandrescu, A. T., Abeygunawardana, C. & Shortle, D. (1994). Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. *Biochemistry*, **33**, 1063-1072.

Alm, E. & Baker, D. (1999a). Matching theory and experiment in protein folding. *Curr. Opin. Struct. Biol.* **9**, 189-196.

Alm, E. & Baker, D. (1999b). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl Acad. Sci. USA*, **96**, 11305-11310.

Blanco, F. J., Ortiz, A. R. & Serrano, L. (1997). ^{1}H and ^{15}N NMR assignment and solution structure of the SH3 domain of spectrin: comparison of unrefined and refined structure sets with the crystal structure. *J. Biomol. NMR*, **9**, 347-353.

Blanco, F. J., Serrano, L. & Forman-Kay, J. D. (1998). High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. *J. Mol. Biol.* **284**, 1153-1164.

Bond, C. J., Wong, K. B., Clarke, J., Fersht, A. R. & Daggett, V. (1997). Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc. Natl Acad. Sci USA*, **94**, 13409-13413.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet.* **21**, 167-195.

Chiti, F., Taddei, N., Webster, P., Hamada, D., Fiaschi, T., Ramponi, G. & Dobson, C. M. (1999a). Acceleration of the folding of acylphosphatase by stabilization of local secondary structure. *Nature Struct. Biol.* **6**, 380-387.

Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999b). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005-1009.

Creighton, T. E. (1995). Protein folding: an unfolding story. *Curr. Biol.* **5**, 353-356.

de Prat, Gay G., Ruiz-Sanz, J., Neira, J. L., Corrales, F. J., Otzen, D. E., Ladurner, A. G. & Fersht, A. R. (1995). Conformational pathway of the polypeptide chain of chymotrypsin inhibitor-2 growing from its N terminus *in vitro*. Parallels with the protein folding pathway. *J. Mol. Biol.* **254**, 968-979.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

Dobson, C. M. & Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92-101.

Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. USA,* **92**, 10869-10873.

Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771-782.

Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B. & Feeney, J. (1990). HMQC-NOESY-HMQC, a three-dimensional NMR experiment which allows detection of nuclear Overhauser effects with protons between overlapping signals. *J. Magn. Reson.* **90**, 420-425.

Galzitskaya, O. V. & Finkelstein, A. V. (1999). A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl Acad. Sci. USA,* **96**, 11299-11304.

Gillespie, J. R. & Shortle, D. (1997a). Characterization of long-range structure in the denatured state of *staphylococcal* nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158-169.

Gillespie, J. R. & Shortle, D. (1997b). Characterization of long-range structure in the denatured state of *staphylococcal* nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**, 170-184.

Hennig, M., Bermel, W., Spencer, A., Dobson, C. M., Smith, L. J. & Schwalbe, H. (1999). Side-chain conformations in an unfolded protein: $\chi 1$ distributions in denatured hen lysozyme determined by heteronuclear ^{13}C, ^{15}N NMR spectroscopy. *J. Mol. Biol.* **288**, 705-723.

Ikura, M., Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). Detection of nuclear Overhauser effects between degenerate amide resonances by heteronuclear three-dimensional nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **112**, 9020-9022.

Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81-91.

Johnson, B. A. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR,* **4**, 603-614.

Kataoka, M. & Goto, Y. (1996). X-ray solution scattering studies of protein folding. *Fold. Des.,* R107-R114.

Mal, T. K., Matthews, S. J., Kovacs, H., Campbell, I. D. & Boyd, J. (1998). Some NMR experiments and a structure determination employing a [^{15}N ^{2}H] enriched protein. *J. Biomol. NMR,* **12**, 259-276.

Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010-1016.

Martinez, J. C., Pisabarro, M. T. & Serrano, L. (1998). Obligatory steps in protein folding and the conformational diversity of the transition state. *Nature Struct. Biol.* **5**, 721-729.

Mok, Y. K., Kay, C. M., Kay, L. E. & Forman-Kay, J. (1999). NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J. Mol. Biol.* **289**, 619-638.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. *Nature,* **359**, 851-855.

Neira, J. L. & Fersht, A. R. (1999). Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. *J. Mol. Biol.* **285**, 1309-1333.

Pachter, R., Arrowsmith, C. H. & Jardetzky, O. (1992). The effect of selective deuteration on magnetization transfer in larger proteins. *J. Biomol. NMR,* **2**, 183-194.

Prieto, J., Wilmans, M., Jimenez, M. A., Rico, M. & Serrano, L. (1997). Non-native local interactions in protein folding and stability: introducing a helical tendency in the all beta-sheet alpha-spectrin SH3 domain. *J. Mol. Biol.* **268**, 760-778.

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.

Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* **6**, 1016-1024.

Shoemaker, B. A. & Wolynes, P. O. (1999). Exploring structures in protein folding funnels with free energy functionals: the denatured ensemble. *J. Mol. biol.* **287**, 657-674.

Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27-34.

Sinclair, J. F. & Shortle, D. (1999). Analysis of long-range interactions in a model denatured state of *staphylococcal* nuclease based on correlated changes in backbone dynamics. *Protein Sci.* **8**, 991-1000.

Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. M. (1996). The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold. Des.* **1**, R95-106.

Sosnick, T. R., Mayne, L. & Englander, S. W. (1996). Molecular collapse: the rate-limiting step in two-state cytochrome c folding. *Proteins: Struct. Funct. Genet.* **24**, 413-426.

Tsai, J., Levitt, M. & Baker, D. (1999). Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol.* **291**, 215-225.

Viguera, A. R., Jimenez, M. A., Rico, M. & Serrano, L. (1996a). Conformational analysis of peptides corresponding to beta-hairpins and a beta-sheet that represent the entire sequence of the alpha-spectrin SH3 domain. *J. Mol. Biol.* **255**, 507-521.

Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L. & Serrano, L. (1994). Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry,* **33**, 2142-2150.

Viguera, A. R., Serrano, L. & Wilmanns, M. (1996b). Different folding transition states may result in the same native structure. *Nature Struct. Biol.* **3**, 874-880.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR,* **6**, 135-140.

Wüthrich, K. (1994). NMR assignments as a basis for structural characterization of denatured states of proteins. *Curr Opin. Struct. Biol.* **4**, 93-99.

Zhang, O. & Forman-Kay, J. D. (1995). Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry,* **34**, 6784-6794.

Zhang, O. & Forman-Kay, J. D. (1997). NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry,* **36**, 3959-3970.

Zhang, O., Forman-Kay, J. D., Shortle, D. & Kay, L. E. (1997a). Triple-resonance NOESY-based experiments

with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. *J. Biomol. NMR*, **9**, 181-200.

Zhang, O., Kay, L. E., Shortle, D. & Forman-Kay, J. D. (1997b). Comprehensive NOE characterization of a partially folded large fragment of *staphylococcal* nuclease Δ131Δ, using NMR methods with improved resolution. *J. Mol. Biol.* **272**, 9-20.

Zhu, G. & Bax, A. (1992). Two-dimensional linear prediction for signals truncated in both dimensions. *J. Magn. Reson.* **98**, 192-199.

Edited by A. R. Fersht

(Received 21 December 1999; accepted 18 February 2000)