

Equilibrium Folding Intermediates of a Greek Key β -Barrel Protein

Stefan Bagby¹, Sandy Go¹, Sumiko Inouye², Mitsuhiro Ikura¹
and Avijit Chakrabarty^{1*}

¹*Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario Canada, M5G 2M9*

²*Department of Biochemistry Robert Wood Johnson Medical School, Piscataway NJ 08854, USA*

*Corresponding author

Protein S is a calcium-binding protein comprising two Greek key β -barrel domains. We have used NMR and optical spectroscopies to show that, in the absence of calcium, the N-terminal domain of protein S forms two equilibrium folding intermediates that are in slow exchange. The intermediates arise from differential calcium-dependent folding of subdomains which are not contiguous along the polypeptide chain. The structures of these intermediates are incompatible with several previously proposed folding mechanisms for Greek key β -barrel domains. We propose a different mechanism that involves multiple nucleation sites for folding and sequential acquisition of native long-range interactions.

© 1998 Academic Press Limited

Keywords: Greek key β -barrel; protein folding intermediates; molten globule; calcium-dependent folding; NMR and optical spectroscopies

Introduction

The Greek key motif was recognized as a super-secondary structure by Richardson in 1977 (Richardson, 1977). It is composed of four antiparallel β -strands with +3, -1, -1 connectivities. In two-dimensional graphical representations, the motif resembles and is named after the repeating unit of a decorative pattern frequently found on Greek pottery (Richardson, 1977). A distinguishing feature of the Greek key is that the first and fourth β -strands are at opposite ends of the primary structure, but are adjacent in the motif. All-Greek key domains invariably comprise a pair of Greek keys. Richardson proposed a feasible pathway for acquiring the long range tertiary interactions found in Greek key proteins whereby the polypeptide chain folds into a long two-stranded " β -ribbon" which then bends in the middle to form the Greek key (Richardson, 1977). A more recently proposed folding mechanism involves a different nucleation structure, the " β -zipper" (Hazes & Hol, 1992). It

has been pointed out, however, that the three-dimensional structures of Greek keys show extensive variability (Hutchinson & Thornton, 1993), so that a universal folding mechanism for Greek keys seems unlikely.

The folding mechanisms of the eye lens $\beta\gamma$ -crystallins (Jaenicke, 1996), and γ II-crystallin in particular (Rudolph *et al.*, 1990), have been studied in some detail. γ II-Crystallin is a two-domain protein in which each domain contains two Greek keys. The γ II-crystallin folding studies strongly suggest that individual domains fold independently in two-state reactions (Rudolph *et al.*, 1990). The experimental data on γ II-crystallin folding are consistent with either the β -ribbon or β -zipper hypothesis and, therefore, cannot be used to validate one model over the other.

Here we report studies on the structure and stability of the N-terminal domain of protein S from *Myxococcus xanthus*. Protein S is a developmentally regulated, Greek key-containing protein which forms a multilayer protective coat on the surface of myxospores and may also function in cell-cell adhesion in a calcium-dependent manner. Protein S resembles the $\beta\gamma$ -crystallins both in sequence (Wistow *et al.*, 1985) and structure (Bagby *et al.*, 1994a,b). Each of the two domains of protein S binds a single calcium ion. The N-terminal domain is stable in isolation (Inouye *et al.*, 1983) as is the N-terminal domain of γ II-crystallin (Rudolph *et al.*, 1990). One differentiating feature of protein S

Abbreviations used: ANS, 1-anilinonaphthalene-8-sulfonic acid; CD, circular dichroism; CSI, chemical shift index; FPLC, fast protein liquid chromatography; HSQC, heteronuclear single quantum correlation; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; protein S-NTD, protein S N-terminal domain; TFE, trifluoroethanol; 2D, two dimensional.

is an α -helix inserted into a loop in each domain (Bagby *et al.*, 1994b).

We find that in the calcium-free state, the isolated N-terminal domain of protein S (residues 1 to 92; abbreviated protein S-NTD) is partially folded. The β -ribbon (Richardson, 1977) and β -zipper (Hazes & Hol, 1992) folding mechanisms both appear to be incompatible with the structure of the partially folded calcium-free molecule. We propose a different folding mechanism for protein S-NTD that is more consistent with the structural data.

Results and Discussion

Structure of calcium-bound protein S N-terminal domain

Protein S consists of two β -barrel domains (residues 1 to 86) and (92 to 173) connected by a short linker (Bagby *et al.*, 1994a). The principal components of the domain architecture are two Greek key motifs, one of which contains an α -helix in addition to the usual four β -strands. The two Greek keys of each domain form two β -sheets, each sheet comprising the first, second and fourth strands from one Greek key and the third strand from the other Greek key. The two sheets lie opposite one another in a wedge or V arrangement. Two rows of side-chains project from each sheet to form the hydrophobic core of each domain.

A protein S-like fold is adopted by the $\beta\gamma$ -crystallins, proteins that are responsible for maintaining a smooth gradient of refractive index in the vertebrate eye lens (Wistow, 1990; Wistow & Piatigorsky, 1988); the $\beta\gamma$ -crystallins have the same β -strand topology as protein S, but lack the two α -helices observed in protein S. A yeast killer toxin from *Williopsis mrakii* adopts the same fold as the $\beta\gamma$ -crystallins (Antuch *et al.*, 1996) and represents a potential single-domain precursor of the protein S/ $\beta\gamma$ -crystallin structure type (Antuch *et al.*, 1996; Editorial, 1996) The existence of this single-domain polypeptide, together with strong evidence that individual domains of γ II-crystallin fold independently (Rudolph *et al.*, 1990), supports the physiological relevance of our studies of the isolated N-terminal domain of protein S. Spherulin 3a from the slime mold *Physarum polycephalum* has been identified by sequence comparisons as another potential single-domain precursor of protein S and the $\beta\gamma$ -crystallins (Wistow, 1990). The three-dimensional structure of spherulin 3a, however, remains unknown.

The protein S domain interface is formed by residues from strands F and G and the FG loop of the N-terminal domain and strand B and the α -helix of the C-terminal domain (see below for strand nomenclature). The aromatic ring of Tyr121 extends from this α -helix to pack against the side-chains of Val65, Ala67, Ala69 and Leu108. The ring hydroxyl group of Tyr121 forms two interdomain hydrogen bonds with the backbone amide groups

of Val66 and Ala67. The domain interface buries approximately 1100 Å² of protein surface area.

We have obtained sequential ¹HN, ¹⁵N, ¹³C α and ¹³C β resonance assignments for both intact protein S (Bagby *et al.*, 1994c) and protein S-NTD (this study; chemical shift data to be submitted to BioMagResBank). The chemical shifts are similar. The average chemical shift changes and root-mean-square deviations (rmsd) for the different nuclei are as follows: ¹HN 0.11(±0.18) ppm, ¹⁵N 0.57(±0.88) ppm, ¹³C α 0.28(±0.45) ppm and ¹³C β 0.36(±0.53) ppm. The secondary chemical shifts of ¹³C α and ¹³C β are consequently similar, indicating that the extent and location of regular secondary structure is essentially the same in protein S-NTD and intact protein S (Bagby *et al.*, 1994a,c).

A detailed structural analysis would be required for rigorous definition of the location and extent of tertiary structure changes between intact protein S and its isolated N-terminal domain. However, the chemical shift differences between corresponding peaks in protein S and protein S-NTD are mostly small, reflecting subtle changes in conformation which may result from variation in sample conditions (in particular, pH and temperature) as much as from the absence of the C-terminal domain. For example, a particularly clear demonstration of retention of native tertiary structure in protein S-NTD is provided by the upfield-shifted amide proton resonances of Gln13 and Ala59 (Figure 1(c)): retention of these characteristic shifts from protein S to protein S-NTD indicates that the so-called folded hairpins that connect the first two strands of each Greek key motif in protein S are preserved in protein S-NTD. The folded hairpins, which bring two aromatic rings into proximity, are a characteristic feature of protein S (Bagby *et al.*, 1994a) and $\beta\gamma$ -crystallin structures (Bax *et al.*, 1990; Blundell *et al.*, 1981; Lapatto *et al.*, 1991; Wistow *et al.*, 1983).

There are two clusters of significant chemical shift differences between protein S and protein S-NTD. One comprises residues 65 to 71 that lie at the interface between domains in the full length protein. The average chemical shift changes and root-mean-square deviations for this segment are ¹HN 0.33(±0.31) ppm, ¹⁵N 2.00(±1.01) ppm, ¹³C α 0.32(±0.30) ppm and ¹³C β 0.94(±0.74) ppm. Removal of the C-terminal domain alters the chemical environment of these residues by eliminating inter-domain packing interactions. The other cluster of significant chemical shift changes is at the C terminus of protein S-NTD, comprising residues 88 to 92. The average chemical shift changes and root-mean-square deviations for this segment are ¹HN 0.38(±0.26) ppm, ¹⁵N 1.01(±1.36) ppm, ¹³C α 1.30(±1.09) ppm and ¹³C β 0.93(±0.61) ppm. In intact protein S, these residues also lie close to the C-terminal domain: they are located beyond the end of the N-terminal domain and form part of the inter-domain linker.

The chemical shift similarities between protein S and protein S-NTD suggest that isolating the

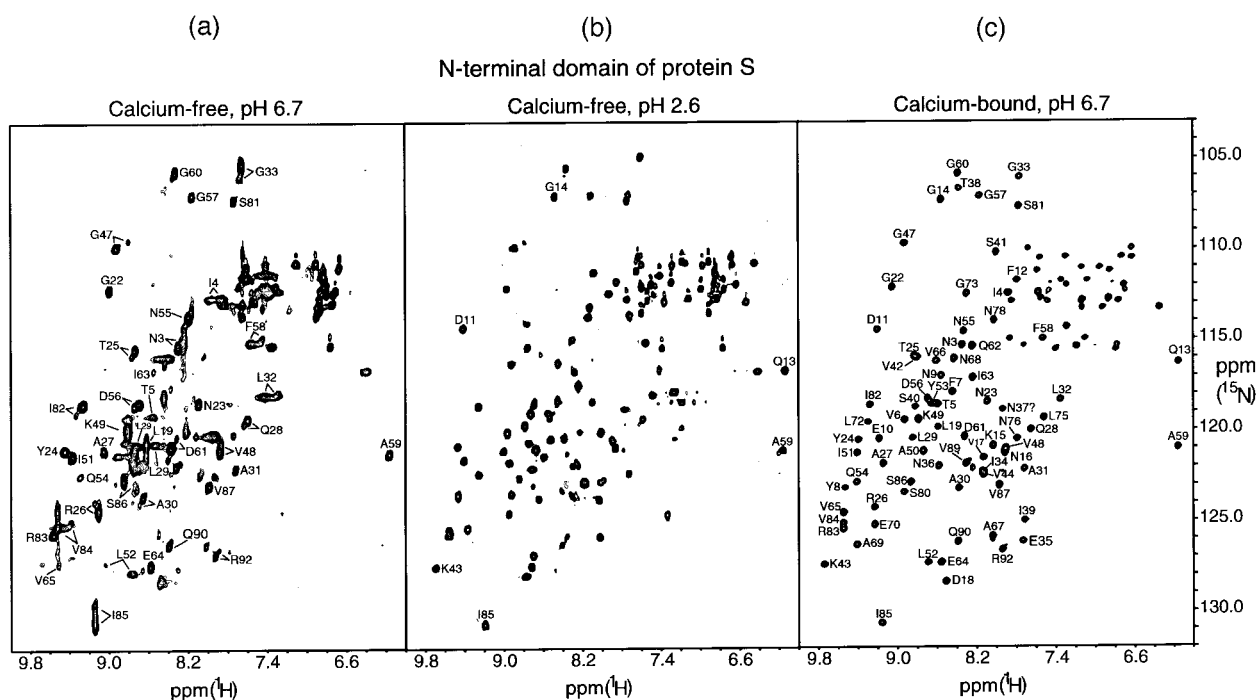


Figure 1. ^1H - ^{15}N HSQC spectra of protein S-NTD. (a) Calcium-free protein S-NTD at pH 6.7; (b) calcium-free protein S-NTD at pH 2.6; (c) calcium-bound protein S-NTD at pH 6.7. For the states of protein S-NTD represented in spectra (a) and (c), cross peaks have been assigned to the backbone NH groups of specific amino acid residues in the protein S-NTD sequence: peaks are labeled according to these assignments using the one-letter amino acid code and sequence position of the amino acid residue. The labeled cross peaks in spectrum (b) are those that are absent in the ^1H - ^{15}N HSQC spectrum of calcium-free protein S-NTD at pH 6.7, but that appear in the ^1H - ^{15}N HSQC spectrum of calcium-free protein S-NTD at pH 2.6, and that can be assigned by inspection by virtue of distinctive chemical shifts.

N-terminal domain and hence removing the domain-domain interactions of the intact protein has little effect on the overall structure of the domain. Only those portions of the N-terminal domain that are close to the C-terminal domain in protein S show significant chemical shift changes in the absence of the C-terminal domain. In further support of close structural similarity between protein S and protein S-NTD, the latter reproduces the former's function of calcium-dependent self-assembly on the myxospore surface (Inouye *et al.*, 1981).

Structure of calcium-free protein S N-terminal domain

The ^1H - ^{15}N HSQC spectrum of calcium-free protein S-NTD at 30°C contains 115 cross-peaks from backbone NH groups, compared with 85 cross-peaks from backbone NH groups for the calcium-bound molecule (Figure 1(a) and (c)). For a single-conformation 92-residue polypeptide with six prolines, ideally we expect 86 cross-peaks from backbone NH groups. The disparity between the number of cross-peaks in the ^1H - ^{15}N HSQC spectrum and the number of backbone NH groups in calcium-free protein S-NTD results from conformational exchange processes such that some backbone NH groups give rise to multiple NH cross-peaks at different chemical shifts. Some of the 115 cross-peaks derive from 42 assignable backbone

NH groups. Lack of sequential assignments for the remaining residues due to apparent absence of signal precludes determination of the three-dimensional structure of calcium-free protein S-NTD by NMR. We can, however, construct a coherent picture of the structure and dynamics of the calcium-free molecule by combining partial resonance assignments with CD spectroscopy, urea denaturation and ANS binding experiments.

The far-UV CD spectra of protein S-NTD both in the presence and absence of calcium contain a minimum at 218 nm and a maximum at approximately 200 nm, indicating the presence of β -structure (Figure 2(a)). The magnitudes of the far-UV CD bands are marginally smaller in the calcium-free molecule. The near-UV CD spectrum of the calcium-bound molecule has five well-resolved bands between 250 and 290 nm which arise from the three phenylalanine and three tyrosine residues of protein S-NTD. In the absence of calcium, the intensity of these bands is diminished by approximately one half (Figure 2(b)). While minor changes in far-UV CD suggest that the secondary structure of protein S-NTD is relatively unaffected by calcium binding, the tertiary structure appears to be stabilized significantly by calcium as reflected in the large changes in near-UV CD.

Calcium-dependent changes in ANS fluorescence were observed when ANS was added to protein S-NTD (Figure 2(c)). Calcium-free protein S-NTD

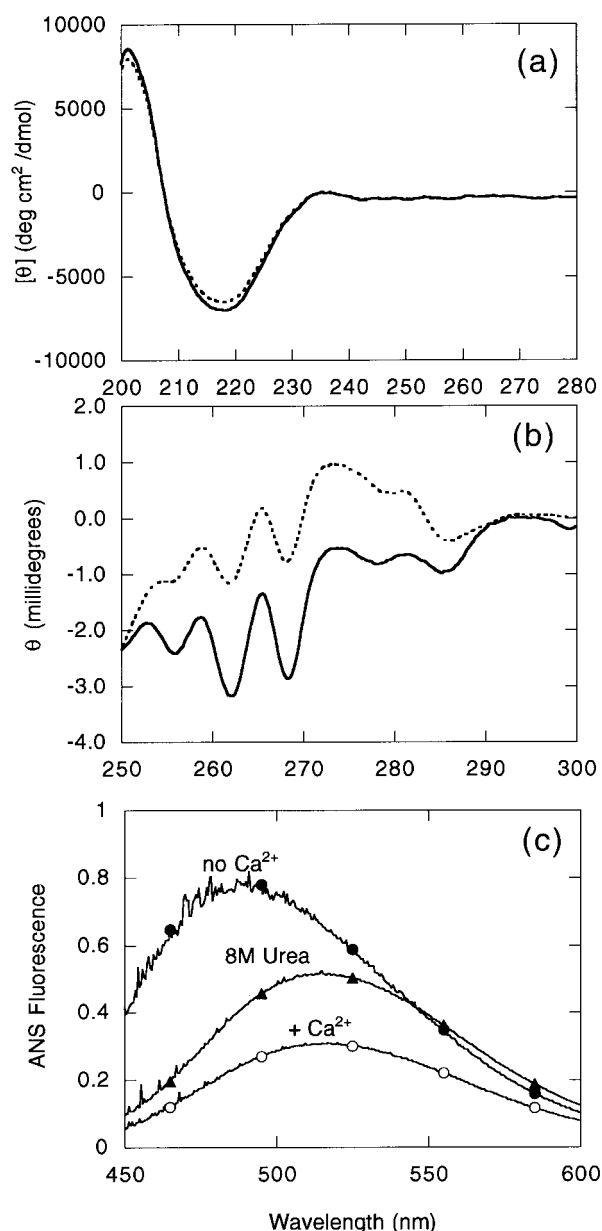


Figure 2. (a) Far-UV CD spectra of calcium-bound protein S-NTD (solid line) and calcium-free protein S-NTD (dashed line). (b) Near-UV CD spectra of calcium-bound protein S-NTD (solid line) and calcium-free protein S-NTD (dashed line). (c), ANS fluorescence spectra in the presence of calcium-bound protein S-NTD (○), calcium-free protein S-NTD (●), and unfolded protein S-NTD in 8.0 M urea (▲).

produced the greatest ANS fluorescence, urea-unfolded protein S-NTD produced intermediate fluorescence, and calcium-bound protein S-NTD under native conditions produced least fluorescence. ANS fluorescence increases when the dye is sequestered in a hydrophobic environment. In the absence of calcium, therefore, protein S-NTD exhibits significant exposure of clustered hydrophobic groups.

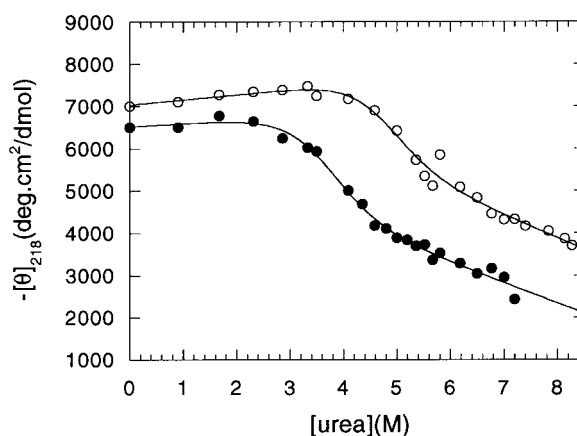


Figure 3. Urea denaturation of calcium-bound (○) and calcium-free protein S-NTD (●). Ellipticity at 218 nm was measured at pH 7.0, 25°C.

The stability of protein S-NTD at pH 7.0, 25°C was assessed from urea denaturation experiments monitored by CD at 218 nm (the CD at 218 nm of protein S-NTD does not change over the pH range 2.5 to 9.0, and temperature range 0 to 80°C; data not shown). Urea denaturation curves of protein S-NTD in the presence and absence of calcium (Figure 3) are atypical and consistent with either two-state unfolding with a sloping post-transition baseline (Santoro & Bolen, 1988) or a non-two-state unfolding process. Calcium binding displaces the denaturation curve to higher ellipticity values throughout the urea concentration range. The similar shapes of the two curves suggest a similar mechanism and cooperativity of unfolding with and without calcium. In the presence of calcium, the curve is also shifted to the right, such that the C_m increases from 3.76 M to 4.86 M. This indicates that calcium ions stabilize protein S-NTD, presumably because the folded state has a greater affinity for calcium than the unfolded state.

Three-dimensional triple-resonance NMR experiments were used to obtain ¹HN, ¹⁵N, ¹³C α and ¹³C β backbone assignments of the detectable resonances of calcium-free protein S-NTD. As indicated on the ¹H-¹⁵N HSQC spectrum (Figure 1), the following segments could be assigned: 3 to 5, 19 to 33, 46 to 65 (with the exception of residues Ala50, Tyr53, and Gln62), and 81 to 92. The chemical shifts of the detectable resonances are similar to those of the corresponding resonances of calcium-bound protein S-NTD: the average chemical shift changes for the different nuclei are ¹HN 0.06(±0.05) ppm, ¹⁵N 0.40(±0.59) ppm, ¹³C α 0.20(±0.20) ppm and ¹³C β 0.32(±0.66) ppm. The values of chemical shift indices (CSI; Wishart & Sykes, 1994a,b) calculated from the assignable resonances of calcium-free protein S-NTD are highly correlated with CSI values from calcium-bound protein S-NTD (Figure 4). The conformations of the segments comprising residues 3 to 5, 19 to 33, 46 to 65 and 81 to 92 in calcium-free protein S-NTD are

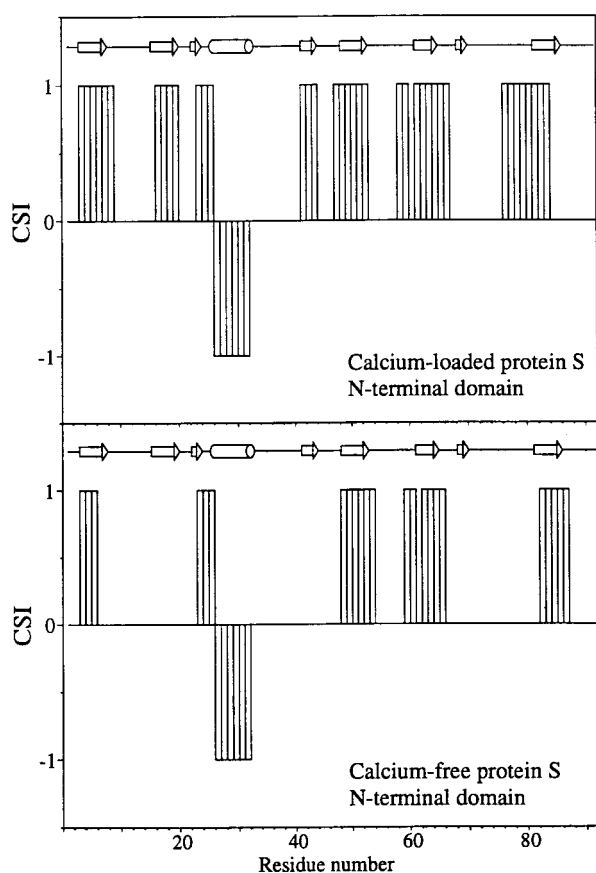


Figure 4. Consensus chemical shift index (CSI; Wishart & Sykes, 1994a,b) versus protein sequence for (a) calcium-bound protein S-NTD and (b) calcium-free protein S-NTD. A CSI value of +1 indicates β -conformation and -1 indicates α -conformation for a particular residue. The locations of secondary structure elements in protein S, determined on the basis of NOE and $^3J_{\text{HN}}$ data (Bagby *et al.*, 1994a), are shown schematically above each CSI plot: arrows represent the location of β -strands and the α -helix is represented by a cylinder.

therefore probably similar to those of the corresponding segments in calcium-bound protein S-NTD.

Residues in the following segments of the protein S-NTD sequence could not be sequentially assigned in the calcium-free molecule due to lack of signal: 1 to 2, 6 to 18, 34 to 45, and 66 to 80. One possibility is that intermediate exchange processes account for the absent cross-peaks, suggesting that these segments of calcium-free protein S-NTD exhibit conformational fluctuations. The first two segments correspond to the N terminus and C-terminal half of strand A, the AB loop, and the N-terminal half of strand B, the third segment corresponds to part of the CD loop and strand D, and the fourth segment corresponds to the FG loop, strand G, and the GH loop. The unassigned segments are in close spatial proximity, constitute most of the BADG β -sheet, and radiate from the putative calcium-ligating residues Glu10 and

Glu71 (Bagby *et al.*, 1994a). A structural model of protein S-NTD, including depiction of the regions most sensitive to calcium binding, is shown in Figure 5.

Instead of chemical exchange broadening due to a folding equilibrium within a monomeric molecule, the absence of cross-peaks could be interpreted in terms of an aggregation equilibrium. Since transient aggregation depends on weak interactions that should be sensitive to protein concentration, the influence of aggregation can be assessed by comparing NMR spectra recorded at low protein concentration with spectra recorded at high protein concentration. The ^1H - ^{15}N HSQC spectrum of calcium-free protein S-NTD recorded using a 90 μM sample (data not shown) displays similar line broadening characteristics and cross-peak absences for the same residues as the ^1H - ^{15}N HSQC spectrum recorded using a 1.5 mM sample. This indicates that transient aggregation is unlikely to contribute significantly to the observed line broadening and cross-peak absences.

Furthermore, the predominant form of both calcium-free and calcium-bound protein S-NTD was found to be monomeric by three different methods. Using pulsed field gradient spin-echo spectroscopy (Dingley *et al.*, 1995; Wu *et al.*, 1995), we determined that the translational diffusion coefficients of 1.5 mM calcium-free protein S-NTD (molecular mass based on amino acid sequence = 9994 Da) and calcium-bound protein S-NTD are approximately 12% and 20% greater than the translational diffusion coefficient of 1.5 mM chicken egg white lysozyme (molecular mass based on amino acid sequence = 14313 Da) and approximately three times smaller than the translational diffusion coefficient of streptomycin sulfate (molecular mass 1457.4 Da). Assuming that chicken egg white lysozyme is predominantly monomeric at 1.5 mM concentration, the larger diffusion coefficients of calcium-free protein S-NTD and calcium-bound protein S-NTD compared with that of lysozyme indicate that both calcium-free protein S-NTD and calcium-bound protein S-NTD are predominantly monomeric under the conditions (1.6 mM protein, pH 6.7, 100 mM KCl) used for acquisition of ^1H - ^{15}N HSQC (except the pH 2.6 sample mentioned later) and triple-resonance NMR data.

As a second test, NMR samples of calcium-free and calcium-bound protein S-NTD at pH 6.7 were injected on to a Superdex-75 HR10/30 FPLC gel filtration column which had been calibrated with protein standards. The apparent molecular mass determined by this method were 12300 Da for calcium-free protein S-NTD and 11200 Da for calcium-bound protein S-NTD. The smaller diffusion coefficient and greater apparent molecular mass of calcium-free protein S-NTD compared with calcium-bound protein S-NTD reflects a difference in hydrodynamic radius, possibly resulting from partial unfolding and the consequently less compact nature of the domain in the absence of calcium. We also monitored the CD at 218 nm of protein



Figure 5. Ribbon representation of the three-dimensional structure of protein S-NTD. The coordinates were taken from a coordinate set for intact protein S (Bagby *et al.*, 1994a), PDB code 1prs, based on the assumption that isolated protein S-NTD has essentially the same structure as in protein S. White indicates those residues for which sequential resonance assignments could not be obtained due to conformational flexibility. Assignable residues in which backbone NH groups exhibit slow two-state exchange are yellow. Assignable residues in which backbone NH groups do not exhibit exchange are green. The bound calcium ion is represented by a cyan sphere. β -strands and the putative calcium ligands Glu10 and Glu71 are labeled.

S-NTD dissolved in 4.2 M urea, within the transition region of the denaturation curves of both calcium-bound and calcium-free protein S-NTD (Figure 3) where the CD is most sensitive to aggregation. The CD was concentration-independent over a 10 to 400 μ M range of protein concentration (data not shown), indicating that both calcium-bound and calcium-free protein S-NTD have low or zero tendency to aggregate.

Residues in the CD (33 to 40) and GH (72 to 80) loops of the N-terminal domain were noted previously to constitute putatively the most flexible portions of this domain in calcium-bound intact protein S (Bagby *et al.*, 1994c). Specifically, the resonance assignments of Asn37, Thr38, Leu72 and Val79 were incomplete. Unassigned cross peaks in the ^1H - ^{15}N HSQC spectrum of protein S, at least some of which must correspond to these unassigned residues, displayed negative or low positive values of the $^{15}\text{N}\{^1\text{H}\}$ NOE, indicative of high flexibility (Bagby *et al.*, 1994c). In calcium-bound protein S-NTD, moreover, the backbone assignments of Asn36, Asn37, Asn77, Asn78 and Val79 were incomplete or lacking altogether. These observations indicate that the more widespread conformational flexibility evidenced in the behavior of calcium-free protein S-NTD is incipient in the structures of calcium-bound protein S and protein S-NTD. In this light, it is pertinent to note another previous observation that approximately 20% of backbone NH cross-peaks are absent from the ^1H - ^{15}N HSQC spectrum of calcium-free protein S relative to the ^1H - ^{15}N HSQC spectrum of calcium-bound protein

S (Bagby *et al.*, 1994b). These and our other observations are consistent with the fact that protein S crystallizes differently in the presence and absence of calcium (Inouye *et al.*, 1980).

Most of the assignable backbone NH groups give rise to two peaks of different intensities in the ^1H - ^{15}N HSQC spectrum of calcium-free protein S-NTD at 30°C (Figure 1(a)), indicating that the NH groups are undergoing slow exchange between two differently populated states. Since the chemical shifts of the doubled cross-peaks are similar to the chemical shifts of the cross-peaks in the ^1H - ^{15}N HSQC spectrum of calcium-bound protein S-NTD (Figure 1(c)), we propose that these peak pairs arise from a two-state exchange involving the native state and a state with near-native structure. The locations in the structure of protein S-NTD of residues undergoing two-state exchange are indicated in Figure 5. The backbone NH groups of the following residues do not display this chemical exchange: Gly22, Asn23, Tyr24, Ala27, Ala31, Lys49, Ile51, Gln54, Gly57, Ala59, Val65, Ser81 and Arg83.

From the results reported above, the structure of calcium-free protein S-NTD can be ascribed the following features. Based on the partial resonance assignments, protein S-NTD can be divided into two subdomains: a calcium-sensitive subdomain (comprising residues 1 to 2, 6 to 18, 34 to 45 and 66 to 80) which contains the putative calcium-ligating residues Glu10 and Glu71 (Bagby *et al.*, 1994a) and undergoes conformational fluctuations in the absence of calcium, and a calcium-insensitive subdomain (composed of residues 3 to 5, 19 to 33, 46

to 65 and 81 to 92). Since the far-UV CD spectrum was relatively unaffected by calcium, the calcium-sensitive subdomain probably retains its β -sheet structure both in the presence and absence of calcium. The calcium-sensitive subdomain also contains three of the six aromatic residues in protein S-NTD (i.e. Phe7, Tyr8 and Phe12) and, thus, the 50% reduction of near-UV CD is probably a result of side-chain mobility of Phe7, Tyr8 and Phe12 in the absence of calcium. The conformational fluctuations of the calcium-sensitive subdomain expose hydrophobic surface area that allows ANS to bind. Residues constituting the calcium-insensitive subdomain can be divided into two groups: those that have a native conformation, and those in two-state exchange between native and near-native conformations. Calcium binding stabilizes the structure of protein S-NTD, but the cooperativity of folding is independent of calcium.

In summary, our structural studies reveal that in the absence of calcium, protein S-NTD forms two well-populated intermediate states that are in slow exchange. The characteristics of the calcium-sensitive subdomain in both intermediates indicate a lack of well-defined structure, and are similar to those of a molten globule (Haynie & Freire, 1993; Kuwajima, 1989). The calcium-insensitive subdomain differs between the two intermediates. In the first intermediate, this subdomain contains a core of residues with native conformation surrounded by residues exchanging between native and near-native conformations. In the second intermediate,

the entire calcium-insensitive subdomain adopts native structure (Figure 6).

Basis of calcium-dependent subdomain folding

Glu10 and Glu71 are two of the residues putatively assigned as calcium ligands in a previous search of the protein S structure for clusters of backbone and side-chain oxygen atoms suitable for calcium ligation (Bagby *et al.*, 1994a). In the absence of calcium, it is conceivable that electrostatic repulsion between negatively charged carboxylate side-chains of these putative calcium-ligating residues contributes to destabilization of the calcium-sensitive subdomain. We investigated this possibility by recording a ^1H - ^{15}N HSQC spectrum of calcium-free protein S-NTD at pH 2.6. Since a typical pK_a of the carboxylate function of glutamate side-chains is 4.4 (Stryer, 1995), these side-chains are almost certainly neutral at pH 2.6 and most electrostatic repulsion should, thus, be removed.

There are 91 backbone NH cross peaks in the ^1H - ^{15}N HSQC spectrum of calcium-free protein S-NTD at pH 2.6 (Figure 1(b)). This corresponds closely to the 86 non-proline residues in protein S-NTD and 85 backbone NH cross-peaks in the ^1H - ^{15}N HSQC spectrum of calcium-bound protein S-NTD at pH 6.7. This numerical similarity and significant reduction in F1 linewidths relative to the pH 6.7 spectrum of calcium-free protein S-NTD suggest that lowering pH removes much of the intermediate time-scale conformational exchange that we deem responsible for the absence of cross-peaks in double and triple-resonance spectra of the calcium-free domain at pH 6.7. In addition, the similar chemical shift dispersion pattern displayed by calcium-free protein S-NTD at pH 2.6 to that displayed by calcium-bound protein S-NTD at pH 6.7 suggests that the calcium-free domain has native-like structure at pH 2.6.

As inferred from the above, most of the protein S-NTD backbone NH cross peaks that are absent from the ^1H - ^{15}N HSQC spectrum of the calcium-free polypeptide are restored at pH 2.6. Obvious examples that we can attribute without sequential assignment of protein S-NTD at pH 2.6, due to their distinctive chemical shifts, are Asp11, Gln13, Gly14 and Lys43. Reappearance of the first three of these peaks close to their chemical shifts in the spectrum of calcium-bound protein S-NTD indicates, for example, that the folded hairpin of the calcium-sensitive subdomain is re-established as a stable, single-conformation motif at pH 2.6. The results of this experiment support the notion that electrostatic repulsion between the side-chains of Glu10 and Glu71 contributes to the destabilization of the calcium-free domain.

In addition to removing conformational exchange, lowering pH by four units must reduce simultaneously the contribution to backbone NH exchange of labile proton exchange with solvent protons; indeed, the intrinsic exchange rate of

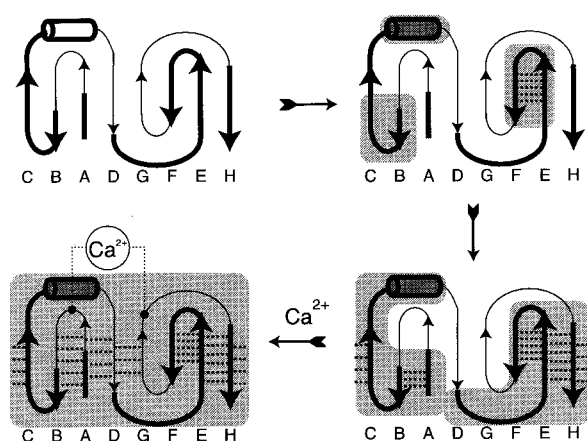


Figure 6. Proposed folding pathway for protein S-NTD. The structure of protein S-NTD is represented as a topology diagram. Regions comprising the calcium-insensitive and calcium-sensitive subdomains are depicted by thick and thin lines. Regions of native structure in folding intermediates are shaded. Hydrogen bonds are indicated by dotted lines. The first intermediate contains the three potential nucleation sites: the tyrosine corner, the α -helix, and the EF β -hairpin. In the second intermediate, the entire calcium-insensitive subdomain possesses native structure. The calcium-ligating residues (Glu10 and Glu71) are represented by dots in the folded calcium-bound structure.

backbone amide protons is a minimum at around pH 2.7 (Connelly *et al.*, 1993). Reduced exchange of labile protons therefore makes a contribution, albeit probably minor, to the observed differences in ^1H - ^{15}N HSQC spectra between pH 6.7 and 2.6.

These results suggest that subdomain unfolding, induced by calcium removal, results from loss of calcium binding energy and from electrostatic repulsion between putative calcium ligands Glu10 and Glu71. By analysis of the NMR spectra of calcium-free protein S-NTD at pH 6.7, we have delineated the approximate extent to which these effects propagate through the polypeptide: the calcium-sensitive region forms a subdomain comprising most of the β -sheet, composed of strands BADG, that lies closer to the protein S-NTD calcium-binding site; the calcium-insensitive region forms a subdomain comprising the FEHC β -sheet and parts of β -strands A and B.

Compatibility of the calcium-free structure of protein S N-terminal domain with the β -ribbon and β -zipper folding mechanisms

Following Richardson's original proposal of the β -ribbon hypothesis (Richardson, 1977), Ptitsyn stated the mechanism explicitly (Ptitsyn, 1981), and also proposed a specific pathway for the folding of γII -crystallin. The high sequence (19% identity and 47% similarity) and structural homology between γII -crystallin and protein S indicates that these proteins share a common folding pathway. Therefore, Ptitsyn's pathway may also be applied to protein S. The pathway involves strands C through H only. Folding is initiated by formation of the β -turn between strands E and F, followed by zippering of the chain into a long two-stranded β -ribbon. Formation of the β -ribbon results in the correct association and hydrogen bonding of strand-pairs E-F, D-G and C-H. The β -ribbon then folds once between strand-pairs E-F and D-G, and then again between strand-pairs D-G and C-H. These two folding steps cause the formation of the β -arch connections between strands C and D, D and E, F and G, and G and H, as well as completing the formation of the FEHC β -sheet. Strands A and B, which are not included in the β -ribbon folding pathway, could fold independently into a β -hairpin and hydrogen bond onto the D-G strand-pair to complete formation of the BADG β -sheet.

The alternative Greek key folding mechanism involves nucleation by the β -zipper (Hazes & Hol, 1992), a motif that was detected by structure analysis in several Greek key β -barrel proteins including γII -crystallin. The β -zipper is composed of two β -strands connected by a short loop forming a β -arch. The two strands, while not connected by hydrogen bonds, contain several interdigitated hydrophobic residues which stabilize the putative nucleus. The β -zipper motif in γII -crystallin and protein S is composed of strands D and E, and the folding pathway involves nucleation through β -zipper formation followed by addition of adja-

cent strands. As strands add on to the nucleus, interstrand hydrogen bond interactions and inter-sheet tertiary interactions form simultaneously.

Our structural studies on protein S-NTD can be used to evaluate the β -ribbon and β -zipper hypotheses. While the calcium-insensitive subdomain of protein S-NTD (strands C, E, F, and H and parts of strands A and B) folds cooperatively into a native-like state in the absence of calcium, the calcium-sensitive subdomain (strands D and G and the remainder of strands A and B) does not form a well-defined structure. It is, thus, unlikely that the calcium-sensitive subdomain nucleates the folding of protein S-NTD in the absence of calcium. The β -ribbon folding pathway requires pairing of strands D and G at an early step. These strands, however, are located in the calcium-sensitive subdomain. Their specific pairing at an early folding step consequently appears improbable. Similarly, the β -zipper folding pathway requires the formation of the β -arch connection between strands D and E. Since strand D is unstable in the calcium-free state, it is unlikely that strand D would contribute to a nucleation site. Both the β -ribbon and β -zipper hypotheses thus appear to be incompatible with the folding of protein S-NTD in the absence of calcium: if protein S-NTD in the absence of calcium were to fold by one of these mechanisms, then the calcium-sensitive subdomain would have to possess stability at an early stage in folding and then become unstable as folding progresses.

Folding of protein S N-terminal domain via multiple nucleation sites

We note that the intermediates characterized in this study are equilibrium intermediates and also note the possibility that these intermediates do not necessarily represent kinetic intermediates on the folding pathway. Equilibrium intermediates with structurally similar kinetic counterparts have, however, been observed for several other proteins including apo-myoglobin (Barrick & Baldwin, 1993), ribonuclease H (Chamberlain *et al.*, 1996) and cytochrome *c* (Bai & Englander, 1996).

If the two equilibrium intermediate states of calcium-free protein S-NTD are similar to kinetic intermediates that accumulate transiently during folding, then the simplest model for folding is a sequential pathway with two intermediates, in which increasing amounts of native structure are acquired at each step (Figure 6). In the first intermediate state, Gly22, Asn23, Tyr24, Ala27, Ala31, Lys49, Ile51, Gln54, Gly57, Ala59, Val65, Ser81 and Arg83 (residues with non-exchanging backbone NH groups) of the calcium-insensitive subdomain acquire their native conformation, the remainder of this subdomain is in a near-native conformation, and the calcium-sensitive subdomain is in a molten globule-like state. Upon conversion to the second intermediate state, the entire calcium-insensitive subdomain adopts native structure and the calcium-sensitive subdomain remains molten-like.

Binding of calcium to the second intermediate induces the calcium-sensitive subdomain to fold, producing the native state of protein S-NTD.

With respect to the folding mechanism of protein S-NTD in the absence of calcium, we hypothesize that the first intermediate involves three nucleation sites formed through local interactions by residues with non-exchanging backbone NH groups. Of the 13 residues that have non-exchanging backbone NH groups, only Ser81 and Arg83 lie outside of these three nucleation sites. The order in which the folding nucleation sites are described below does not reflect any known order of importance or chronology of folding events.

The first nucleus involves α -helix formation in residues 26 to 32. Residues Ala27 and Ala31 of this α -helix have non-exchanging backbone NH groups. Furthermore, analysis of the amino acid sequence of protein S-NTD using helix-coil transition theory (Doig *et al.*, 1994) and experimentally determined helix propensity parameters (Chakrabartty *et al.*, 1994; Rohl *et al.*, 1996) indicates that only residues 26 to 32 possess significant helix propensity. This segment should therefore form a stable helix (estimated fractional helix content of residues 26 to 32 \sim 10%) in the unfolded state; this region corresponds exactly to the single helix in protein S-NTD.

β -hairpin formation between strands E and F (residues 48 to 65) establishes the second nucleus. In the EF β -hairpin, backbone NH groups of residues Lys49, Ile51, Gln54, Gly57, Ala59 and Val65 do not show exchange, evidencing a single, stable conformation. Previous demonstrations that small peptides can form stable β -hairpins in aqueous solution (Blanco *et al.*, 1994; Searle *et al.*, 1995; Sieber & Moe, 1996) provide indirect evidence that the EF β -hairpin serves as a folding nucleus. The EF β -hairpin is one of four so-called folded hairpins which are a prominent feature of the protein S and $\beta\gamma$ -crystallin structures (Bagby *et al.*, 1994a; Lapatto *et al.*, 1991). In calcium-free protein S-NTD, the EF β -hairpin may be stabilized by two interactions that are characteristic of the folded hairpins: the hydrophobic interaction between the aromatic side-chains of Phe58 (in the hairpin) and Tyr53 (at the C-terminal end of strand F); and a hydrogen bond between the backbone amide of Phe58 and the side-chain hydroxyl of Ser81 at the N-terminal end of strand H. The corresponding β -hairpin in the calcium-sensitive subdomain is formed by strands A and B. Glu10, proposed to ligate calcium, is part of this hairpin. We surmise that the AB β -hairpin does not nucleate folding in the absence of calcium because the hydrophobic (between Phe12 and Phe7) and hydrogen bonding (between Phe12 and Ser41) interactions that stabilize this folded hairpin are outweighed by destabilization due to electrostatic repulsion between the negatively charged carboxylate side-chains of the calcium ligands Glu10 and Glu71.

The third nucleation site for folding involves residues 19 to 24, again part of the calcium-insensitive subdomain. Of these six residues, two are pro-

lines and Gly22, Asn23 and Tyr24 have non-exchanging backbone NH peaks. The sequence of this segment, LPPGNY, conforms to the consensus sequence LxPGxY of the tyrosine corner, a motif that is a "nearly ubiquitous feature of low-twist Greek key β -barrel proteins" (Hemmingsen *et al.*, 1994) such as γ II-crystallin and protein S. In this motif, the side-chain hydroxyl of a tyrosine near the beginning or end of an antiparallel β -strand forms a hydrogen bond to the backbone of a residue that precedes the tyrosine by 3 to 5 residues (i.e. Y-3, Y-4, or Y-5). The Leu residue in the Y-5 position packs next to the tyrosine ring, further stabilizing the motif. The position of the tyrosine corner in the structure of protein S is typical of other tyrosine corners. The medium resolution of the protein S structure (Bagby *et al.*, 1994a), however, does not allow unambiguous identification of the tyrosine side-chain-backbone hydrogen bond. If the tyrosine corner possesses intrinsic stability then it may function as a folding initiation site in protein S-NTD. The ubiquity of the tyrosine corner in Greek key proteins with different functions suggests that this motif is involved in the folding mechanism of Greek keys. The location of the tyrosine corner in a structured region of the first folding intermediate with non-exchanging backbone NH groups further supports the tyrosine corner's involvement in the folding mechanism.

In summary, we hypothesize that the calcium-insensitive subdomain of protein S-NTD folds *via* the formation of three nucleation sites (α -helix, EF β -hairpin, and tyrosine corner) which are stabilized by local interactions that persist in the native structure. These sites may form independently or sequentially. Interaction between the sites stabilizes the native-like structure of the calcium-insensitive subdomain. Folding of the calcium-dependent subdomain requires calcium binding, which stabilizes the subdomain and reduces charge repulsion between carboxylates of calcium-ligating residues. While our proposed folding pathway is specific for protein S-NTD, the β -hairpin and tyrosine corner could serve as folding nuclei in other Greek key β -barrel proteins.

Comparison with studies on other β -sheet proteins

The folding mechanism of apoplastocyanin, a Greek key-containing copper-binding protein, has two kinetic phases, both involving *trans-to-cis* isomerization of two prolines (Koide *et al.*, 1993). Amide proton exchange experiments demonstrate formation of a kinetic intermediate, in the very slow refolding phase, with weak protection of amide protons. This intermediate possesses characteristics similar to those we propose for calcium-free protein S-NTD, including flexible β -sheets, altered packing of the hydrophobic core, and localized tertiary interactions.

In line with γ II-crystallin (Rudolph *et al.*, 1990), it is likely that the domains of protein S fold indepen-

dently with the more stable N-terminal domain folding faster than the C-terminal domain. Furthermore, since each domain of both proteins contains a tyrosine corner (Hemmingsen *et al.*, 1994) and EF β -hairpin (Bagby *et al.*, 1994b), these elements are probable nucleation sites along the folding pathways of both N and C-terminal domains of both proteins.

Studies to date indicate that β -sheet proteins can fold through either hierarchical or non-hierarchical pathways. Intermediates along hierarchical pathways contain solely native interactions, whereas intermediates along non-hierarchical pathways contain non-native α -helical elements. Protein S-NTD appears to follow a hierarchical pathway. Interleukin-1 β (Varley *et al.*, 1993) and rat intestinal fatty acid binding protein (Ropson *et al.*, 1990) also belong to this class of β -sheet proteins. In addition, the presence of only native long range interactions in folding intermediates of protein S-NTD resembles the TFE-induced intermediate state of the β -sheet protein tendamistat (Schonbrunner *et al.*, 1997). Non-hierarchical folding pathways have been proposed for β -lactoglobulin (Shiraki *et al.*, 1995) and cellular retinoic acid binding protein (Liu *et al.*, 1994). The isolated N-terminal Src homology 3 (SH3) domain of the *Drosophila* protein drk (Farrow *et al.*, 1995, 1997; Zhang & Forman-Kay, 1995, 1997; Zhang *et al.*, 1994) and mammalian GRB2 (Goudreau *et al.*, 1994) are in equilibrium between folded and unfolded states in aqueous buffer conditions. The structure and dynamics of the folded and unfolded states of the drk SH3 domain have been characterized in detail, permitting comparison with the denatured state in 2 M guanidine hydrochloride. Both unfolded states sample the α -region of ϕ, ψ space suggesting that turn-like structures that do not persist in the predominantly β -sheet native form may play a role in initiation of SH3 domain folding. Unlike the non-hierarchical helical intermediate of β -lactoglobulin, the non-native turn-like structures in the unfolded state of drk N-terminal SH3 domain do not constitute cooperative helical segments. Further studies will determine which pathway, hierarchical or non-hierarchical, is utilized by the majority of β -sheet proteins, and may allow prediction of a hierarchical or non-hierarchical pathway based on amino acid sequence.

Materials and Methods

Protein expression, purification, and NMR sample preparation

Protein S-NTD was overexpressed in *Escherichia coli* using a derivative of the pET11a plasmid carrying a phage T7 promoter (Studier *et al.*, 1990). The protein was purified by $(\text{NH}_4)_2\text{SO}_4$ precipitation followed by ion exchange chromatography using PAE1000 (Amicon) and Mono S (Pharmacia). Sample homogeneity exceeded 95% as judged by SDS-PAGE and 2D NMR spectroscopy. Uniformly ^{15}N and $^{15}\text{N}/^{13}\text{C}$ -labeled protein was obtained by using $^{15}\text{NH}_4\text{Cl}$ and $^{15}\text{NH}_4\text{Cl}/[^{13}\text{C}_6]\text{-D-glucose}$ as the sole nitrogen/carbon sources in M9 medium.

Uniformly ^{15}N -labeled or $^{15}\text{N}/^{13}\text{C}$ -labeled protein was dissolved in 95% $\text{H}_2\text{O}/5\%$ $^2\text{H}_2\text{O}$. NMR samples of calcium-bound protein S-NTD were adjusted to pH 6.7 without correction for isotope effects and contained 1.6 mM protein S-NTD, 10 mM CaCl_2 , 100 mM KCl, and 0.05 mM NaN_3 . In order to make calcium-free protein S-NTD, desalted, lyophilized protein was dissolved over a period of one hour in the minimum possible volume of 20 mM sodium phosphate/2 mM EDTA, pH 8.0. The EDTA was subsequently removed by Sephadex G-25 chromatography. NMR samples were prepared in 100 mM KCl and 0.05 mM NaN_3 , pre-treated with Chelex 100 resin (Bio-Rad).

Circular dichroism spectroscopy

CD measurements were made on an Aviv 62DS spectropolarimeter. Far-UV CD spectra were recorded in a 0.1 cm quartz cell in 1 mM each of sodium borate and sodium phosphate, 25°C, pH 7.0, and samples contained either 10 mM CaCl_2 or 1 mM EDTA. Near-UV CD spectra were recorded under the same conditions in a 1.0 cm quartz cell. Measurements at 218 nm were made in a 1.0 cm quartz cell in 1 mM each of sodium borate and sodium phosphate, 25°C, pH 7.0. The concentration of urea varied between 0 and 8 M, and samples contained either 10 mM CaCl_2 or 1 mM EDTA. CD measurements are reported as mean residue ellipticity at a given wavelength. Protein concentration in stock solutions was determined by dissolving aliquots in 6 M guanidine-HCl and measuring tyrosine absorbance at 275 nm, using $1450 \text{ M}^{-1}\text{cm}^{-1}$ as the extinction coefficient for a single tyrosine (Brandts & Kaplan, 1973).

Fluorescence spectroscopy

Steady-state fluorescence was measured at room temperature using a Photon Technology International QM-1 fluorescence spectrophotometer equipped with excitation intensity correction and a magnetic stirrer. Emission spectra from 450 to 600 nm were collected ($\lambda_{\text{ex}} = 372 \text{ nm}$, 1 second/nm, bandpass = 4 nm for excitation and emission). Samples contained 2 μM protein S-NTD, 2 μM ANS, 1 mM sodium phosphate, 1 mM sodium borate (pH 7.0), and 20 mM CaCl_2 or 20 mM EDTA or 8 M urea.

NMR data acquisition

Unless otherwise stated, NMR spectra were acquired at 30°C on a Varian UNITY-*plus* 500 spectrometer operating at a proton frequency of 499.843 MHz and equipped with an actively shielded Z-gradient and a gradient amplifier unit. The following carrier positions were used: ^{13}C , 176 ppm; ^{15}N , 117.5 ppm; $^{13}\text{C}^\alpha$, 58 ppm; $^{13}\text{C}^\beta$, 43 ppm; ^1H , 4.725 ppm (on resonance with the water frequency). When possible, NMR data were acquired using sensitivity enhancement (Cavanagh & Rance, 1990; Kay *et al.*, 1992) and pulsed-field gradients for coherence transfer pathway selection and artifact suppression (Kay, 1995; Keeler *et al.*, 1994). Sensitivity was further improved by minimizing the degree of water saturation (Grzesiek & Bax, 1993; Kay *et al.*, 1994; Li & Montelione, 1993; Stonehouse *et al.*, 1994).

Two-dimensional ^1H - ^{15}N HSQC spectra (Bodenhausen & Ruben, 1979; Zhang *et al.*, 1994) (Figure 1) were acquired with 128 complex points in t_1 , 1024 complex points in t_2 and 16 transients. The ^1H - ^{15}N HSQC spec-

trum of the 90 μ M sample of calcium-free protein S-NTD was acquired with 170 complex points in t_1 , 1024 complex points in t_2 and 220 transients. Triple-resonance spectra were recorded on uniformly $^{13}\text{C}/^{15}\text{N}$ -labeled protein S-NTD with the following numbers of complex points and spectral widths: HNCB (Kay *et al.*, 1990, 1994; Muhandiram & Kay, 1994) ^{15}N (t_1) 32, 1464 Hz, ^{13}C (t_2) 64, 1500 Hz, ^1H (t_3) 1024, 8000 Hz (24 transients); HNCACB (Muhandiram & Kay, 1994; Wittekind & Mueller, 1993) $^{13}\text{C}^{\alpha/\beta}$ (t_1) 48, 7625 Hz, ^{15}N (t_2) 28, 1464 Hz, ^1H (t_3) 1024, 8000 Hz (48 transients); CBCA-(CO)NH (Grzesiek & Bax, 1992; Muhandiram & Kay, 1994) $^{13}\text{C}^{\alpha/\beta}$ (t_1) 48, 7625 Hz, ^{15}N (t_2) 28, 1464 Hz, ^1H (t_3) 1024, 8000 Hz (24 transients); (HB)CBCACO(CA)HA (Kay, 1993) $^{13}\text{C}^{\alpha/\beta}$ (t_1) 54, 7625 Hz, ^{13}C (t_2) 64, 1500 Hz, ^1H (t_3) 768, 6000 Hz (16 transients); HBHA-(CBCACO)NH (L. E. Kay, personal communication) $^1\text{H}\alpha/\beta$ (t_1) 50, 3000 Hz, ^{15}N (t_2) 32, 1464 Hz, ^1H (t_3) 1024, 8000 Hz (40 transients).

Translational diffusion coefficients were determined from data acquired using a bipolar pulse pair longitudinal-eddy-current delay pulse sequence (Wu *et al.*, 1995). These experiments were performed at 30°C on a Varian UNITY 600 spectrometer using a probe with a 70 G cm^{-1} gradient coil. For streptomycin sulfate (purchased from Sigma and used without further purification), 14 one-dimensional data sets were acquired using 48 transients and gradient pulse durations ranging from 0 to 1.3 ms. For calcium-free protein S-NTD, calcium-bound protein S-NTD and lysozyme (purchased from Sigma and used without further purification), 12 one-dimensional data sets were acquired using 80 transients and gradient pulse durations ranging from 0 to 2.4 ms.

NMR data processing and analysis

All spectra were processed and displayed using NMRPipe and NMRDraw (Delaglio *et al.*, 1995). A 60° phase-shifted sine bell and single zero fill were typically applied prior to Fourier transformation in all dimensions of triple-resonance spectra. In the t_1 dimension of ^1H - ^{15}N HSQC spectra, a 45° phase-shifted sine bell was applied prior to Fourier transformation, and a Lorentzian-to-Gaussian filter with NMRPipe parameters $G1 = 10$ Hz (inverse exponential width), $G2 = 20$ Hz (Gaussian width) and $G3 = 0$ (corresponding to a maximum at the first point in the FID) was applied in the t_2 dimension. For the HNCACB, CBCA(CO)NH, HBHA(CBCACO)NH and (HB)CBCACO(CA)HA data sets, linear prediction (Barkhuijsen *et al.*, 1985) was used to double the number of planes in the t_2 dimension. Where necessary, the residual water signal was reduced using a time-domain deconvolution approach (Marion *et al.*, 1989). The programs CAPP and PIPP (Garrett *et al.*, 1991) were used for peak picking and spectral analysis.

Sequential resonance assignments for calcium-bound protein S-NTD were made using a combination of HNCACB and CBCA(CO)NH spectra, while those for calcium-free protein S-NTD were made using a combination of HNCACB, CBCA(CO)NH, (HB)CBCACO(CA)HA, HBHA(CBCACO)NH and HNCB spectra.

for helpful discussions, Frank Delaglio for supplying the programs NMRPipe and NMRDraw, Dan Garrett for providing the programs CAPP and PIPP, and Rieko Ishima for assistance with diffusion coefficient determination. This work was supported in part by a grant to A.C. from the Medical Research Council of Canada (MRCC), and by a grant to M.I. from the National Cancer Institute of Canada. S.B. thanks the MRCC for the award of a post-doctoral fellowship. M.I. is the holder of an MRCC Scholarship and a Howard Hughes Medical Institute International Scholarship.

References

- Antuch, W., Guntert, P. & Wuthrich, K. (1996). Ancestral $\beta\gamma$ -crystallin precursor structure in a yeast killer toxin. *Nature Struct. Biol.* **3**(8), 662–665.
- Bagby, S., Harvey, T. S., Eagle, S. G., Inouye, S. & Ikura, M. (1994). aNMR-derived three-dimensional solution structure of protein S complexed with calcium. *Structure*, **2**(2), 107–122.
- Bagby, S., Harvey, T. S., Eagle, S. G., Inouye, S. & Ikura, M. (1994b). Structural similarity of a developmentally regulated bacterial spore coat protein to $\beta\gamma$ -crystallins of the vertebrate eye lens. *Proc. Natl Acad. Sci. USA*, **91**, 4308–4312.
- Bagby, S., Harvey, T. S., Kay, L. E., Eagle, S. G., Inouye, S. & Ikura, M. (1994c). Unusual helix-containing Greek keys in development-specific Ca^{2+} -binding protein S. ^1H , ^{15}N , and ^{13}C assignments and secondary structure determined with the use of multidimensional double and triple resonance heteronuclear NMR spectroscopy. *Biochemistry*, **33**(9), 2409–2421.
- Bai, Y. & Englander, S. W. (1996). Future directions in folding: the multi-state nature of protein structure. *Proteins: Struct. Funct. Genet.* **24**(2), 145–151.
- Barkhuijsen, H., De Beer, R., Bovee, W. M. M. J. & Van Ormondt, D. (1985). Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure. *J. Magn. Reson.* **61**, 465–481.
- Barrick, D. & Baldwin, R. L. (1993). Stein and Moore Award address. The molten globule intermediate of apomyoglobin and the process of protein folding. *Protein Sci.* **2**(6), 869–876.
- Bax, B., Lapatto, R., Nalini, V., Driessen, H., Lindley, P. F., Mahadevan, D., Blundell, T. L. & Slingsby, C. (1990). X-ray analysis of βB2 -crystallin and evolution of oligomeric lens proteins. *Nature*, **347**(6295), 776–780.
- Blanco, F. J., Rivas, G. & Serrano, L. (1994). A short linear peptide that folds into a native stable β -hairpin. *Nature Struct. Biol.* **1**(9), 584–590.
- Blundell, T., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I., Turnell, B. & Wistow, G. (1981). The molecular structure and stability of the eye lens: X-ray analysis of γ -crystallin II. *Nature*, **289**, 771–777.
- Bodenhausen, G. & Ruben, D. J. (1979). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**(1), 185–189.
- Brandts, J. R. & Kaplan, K. J. (1973). Derivative spectroscopy applied to tyrosyl chromophores. Studies on ribonuclease, lima bean inhibitor, and pancreatic trypsin inhibitor. *Biochemistry*, **12**, 2011–2024.

Acknowledgements

We thank Lewis Kay, University of Toronto, and members of his group for providing pulse sequences and

- Cavanagh, J. & Rance, M. (1990). Sensitivity improvement in isotropic mixing (TOCSY) experiments. *J. Magn. Reson.* **88**, 72–85.
- Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**(5), 843–852.
- Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nature Struct. Biol.* **3**(9), 782–787.
- Connelly, G. P., Bai, Y., Jeng, M. F. & Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 87–92.
- 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**(3), 277–293.
- Dingley, A. J., Mackay, J. P., Chapman, B. E., Morris, M. B., Kuchel, P. W., Hambly, B. D. & King, G. F. (1995). Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2. *J. Biomol. NMR*, **6**, 321–328.
- Doig, A. J., Chakrabartty, A., Klingler, T. M. & Baldwin, R. L. (1994). Determination of free energies of N-capping in α -helices by modification of the Lifson-Roig theory to include N- and C-capping. *Biochemistry*, **33**(11), 3396–3403.
- Editorial. (1996). $\beta\gamma$ -crystallin redux. *Nature Struct. Biol.* **3**(8), 651–652.
- Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1995). Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry*, **34**(3), 868–878.
- Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1997). Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry*, **36**, 2390–2402.
- Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A common sense approach to peak picking in two-, three, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220.
- Goudreau, N., Cornille, F., Duchesne, M., Parker, F., Tocque, B., Garbay, C. & Roques, B. P. (1994). NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline-rich peptide from Sos. *Nature Struct. Biol.* **1**, 898–907.
- Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293.
- Grzesiek, S. & Bax, A. (1993). The importance of not saturating H₂O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594.
- Haynie, D. T. & Freire, E. (1993). Structural energetics of the molten globule state. *Proteins: Struct. Funct. Genet.* **16**, 115–140.
- Hazes, B. & Hol, W. G. (1992). Comparison of the hemocyanin β -barrel with other Greek key β -barrels: possible importance of the “ β -zipper” in protein structure and folding. *Proteins: Struct. Funct. Genet.* **12**(3), 278–298.
- Hemmingsen, J. M., Gernert, K. M., Richardson, J. S. & Richardson, D. C. (1994). The tyrosine corner: a feature of most Greek key β -barrel proteins. *Protein Sci.* **3**(11), 1927–1937.
- Hutchinson, E. G. & Thornton, J. M. (1993). The Greek key motif: extraction, classification and analysis. *Protein Eng.* **6**(3), 233–245.
- Inouye, S., Inouye, M., McKeever, M. & Sarma, R. (1980). Preliminary crystallographic data for protein S, a development-specific protein of *Myxococcus xanthus*. *J. Biol. Chem.* **255**, 3713–3714.
- Inouye, S., Harada, W., Zusman, D. & Inouye, M. (1981). Development-specific protein S of *Myxococcus xanthus*: purification and characterization. *J. Bacteriol.* **148**, 678–683.
- Inouye, S., Franceschini, T. & Inouye, M. (1983). Structural similarities between the development-specific protein S from a Gram-negative bacterium, *Myxococcus xanthus*, and calmodulin. *Proc. Natl Acad. Sci. USA*, **80**, 6829–6833.
- Jaenicke, R. (1996). Stability and folding of ultrastable proteins: eye lens crystallins. *FASEB J.* **10**(1), 84–92.
- Kay, L. E. (1993). Pulsed-field gradient-enhanced three-dimensional NMR experiment for correlating ¹³C α / β , ¹³C', and ¹H α chemical shifts in uniformly ¹³C-labeled proteins dissolved in H₂O. *J. Am. Chem. Soc.* **115**, 2055–2057.
- Kay, L. E. (1995). Field gradient techniques in NMR spectroscopy. *Curr. Opin. Struct. Biol.* **5**(5), 674–681.
- Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 496–514.
- Kay, L. E., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665.
- Kay, L. E., Xu, G. Y. & Yamazaki, T. (1994). Enhanced-sensitivity triple-resonance spectroscopy with minimal H₂O saturation. *J. Magn. Reson. A*, **109**, 129–133.
- Keeler, J., Clowes, R. T., Davis, A. L. & Laue, E. D. (1994). Pulsed-field gradients: theory and practice. *Methods Enzymol.* **239**, 145–207.
- Koide, S., Dyson, H. J. & Wright, P. E. (1993). Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry*, **32**(46), 12299–310.
- Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular protein structure. *Proteins: Struct. Funct. Genet.* **6**, 87–103.
- Lapatto, R., Nalini, V., Bax, B., Driessen, H., Lindley, P. F., Blundell, T. L. & Slingsby, C. (1991). High resolution structure of an oligomeric eye lens β -crystallin. Loops, arches, linkers and interfaces in β B2 dimer compared to a monomeric γ -crystallin. *J. Mol. Biol.* **222**(4), 1067–1083.
- Li, Y.-C. & Montelione, G. T. (1993). Solvent saturation-transfer effects in pulsed-field-gradient heteronuclear single-quantum-coherence (PFG-HSQC) spectra of polypeptides and proteins. *J. Magn. Reson. B*, **101**, 315–319.
- Liu, Z.-P., Rizo, J. & Gierasch, L. M. (1994). Equilibrium folding studies of cellular retinoic acid binding protein, a predominantly β -sheet protein. *Biochemistry*, **33**, 134–142.
- Marion, D., Ikura, M. & Bax, A. (1989). Improved solvent suppression in one- and two-dimensional

- NMR spectra by convolution of time-domain data. *J. Magn. Reson.* **84**, 425–430.
- Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B*, **103**, 203–216.
- Ptitsyn, O. B. (1981). Protein folding: general physical model. *FEBS Letters*, **131**, 197–202.
- Richardson, J. S. (1977). β -Sheet topology and the relatedness of proteins (review). *Nature*, **495**(500), 268 (5620).
- Rohl, C. A., Chakrabartty, A. & Baldwin, R. L. (1996). Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**, 2623–2637.
- Ropson, I. J., Gordon, J. I. & Frieden, C. (1990). Folding of a predominantly β -structure protein: rat intestinal fatty acid binding protein. *Biochemistry*, **29**, 9591–9599.
- Rudolph, R., Siebendritt, R., Nessler, G., Sharma, A. K. & Jaenicke, R. (1990). Folding of an all- β protein: independent domain folding in γ II-crystallin from calf eye lens. *Proc. Natl Acad. Sci. USA*, **87**, 4625–4629.
- Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α -chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
- Schonbrunner, N., Wey, J., Engels, J., Georg, H. & Keifhaber, T. (1997). Native-like β -structure in a trifluoroethanol-induced partially folded state of the all β -sheet protein tendamistat. *J. Mol. Biol.* **260**, 432–445.
- Searle, M. S., Williams, D. H. & Packman, L. C. (1995). A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-soluble non-native β -hairpin. *Nature Struct. Biol.* **2**(11), 999–1006.
- Shiraki, K., Nishikawa, K. & Goto, Y. (1995). Trifluoroethanol-induced stabilization of the α -helical structure of β -lactoglobulin: implications for non-hierarchical protein folding. *J. Mol. Biol.* **245**, 180–194.
- Sieber, V. & Moe, G. R. (1996). Interactions contributing to the formation of a β -hairpin-like structure in a small peptide. *Biochemistry*, **35**(1), 181–188.
- Stonehouse, J., Shaw, G. L., Keeler, J. & Laue, E. D. (1994). Minimizing sensitivity losses in gradient-selected ^{15}N - ^1H HSQC spectra of proteins. *J. Magn. Reson. A*, **107**, 178–184.
- Stryer, L. (1995). *Biochemistry*, 4th edit., Freeman, New York.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89.
- Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H. & Clore, G. M. (1993). Kinetics of folding of the all β -sheet protein interleukin-1 β . *Science*, **260**, 1110–1113.
- Wishart, D. S. & Sykes, B. D. (1994a). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**(2), 171–180.
- Wishart, D. S. & Sykes, B. D. (1994b). Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363–392.
- Wistow, G. (1990). Evolution of a protein superfamily: relationships between vertebrate lens crystallins and microorganism dormancy proteins. *J. Mol. Evol.* **30** (2), 140–145.
- Wistow, G. J. & Piatigorsky, J. (1988). Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu. Rev. Biochem.* **57**, 479–504.
- Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. & Blundell, T. (1983). X-ray analysis of the eye lens protein γ II-crystallin at 1.9 Å resolution. *J. Mol. Biol.* **170**, 175–202.
- Wistow, G., Summers, L. & Blundell, T. (1985). *Myxococcus xanthus* spore coat protein S may have a similar structure to vertebrate lens $\beta\gamma$ -crystallins. *Nature*, **315**, 771–773.
- Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α - and β -carbon resonances in proteins. *J. Magn. Reson. B*, **101**, 201–205.
- Wu, D., Chen, A. & Johnson, C. S., Jr (1995). An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *J. Magn. Reson. A*, **115**, 260–264.
- 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**(20), 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., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994). Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR*, **4**(6), 845–858.

Edited by P. E. Wright

(Received 24 June 1997; received in revised form 24 November 1997; accepted 25 November 1997)