NMR Studies of Unfolded States of an SH3 Domain in Aqueous Solution and Denaturing Conditions†

Ouwen Zhang‡,§ and Julie D. Forman-Kay*,‡

Biochemistry Research Division, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8, Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8, and Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6

Received November 6, 1996; Revised Manuscript Received January 10, 1997⊗

ABSTRACT: The isolated N-terminal SH3 domain of the *Drosophila* adapter protein drk (drkN SH3 domain) exists in a dynamic equilibrium between a folded (F_{exch}) and an unfolded (U_{exch}) state under native-like buffer conditions [Zhang, O., & Forman-Kay, J. D. (1995) *Biochemistry 34*, 6784−6794]. The effect of binding a proline-rich peptide derived from the protein Sos, a biological target of the drkN SH3 domain, on this equilibrium has been investigated. The stabilization of the F_{exch} folded state upon binding provides an example of the link between binding and protein folding or stabilization. We have compared NMR parameters of the U_{exch} state with those of a denatured state in 2 M guanidine hydrochloride (U_{Gdn}). Variable-temperature experiments demonstrate that interactions in a region encompassing residues Gln 23−Leu 28 in the U_{exch} state are destabilized upon addition of guanidine hydrochloride. Data from an ^{15}N HSQC−NOESY−HSQC experiment as well as recently developed methods provide more unambiguous structural information than described previously, showing the presence of preferential structure in both unfolded states. Backbone NOEs observed in both unfolded states as well as chemical shifts and coupling constants suggest a rapid equilibrium between extended structure and turn-like structures which may play a role in initiation of protein folding. However, differences in detailed structural features between the two unfolded states argue that caution is needed in interpretation of results from structural characterization of protein conformational states generated using denaturing conditions.

Conformational changes of proteins from disordered to ordered states occur in many biochemical processes both *in vitro* and *in vivo*, including protein folding and interactions with other proteins, nucleic acids, metals, or ligands (Creighton, 1992; Wilson & Stanfield, 1994; Spolar & Record, 1994; Berg & Shi, 1996). Some of these structural changes are global (folding), while others involve small, local rearrangements. Unfolded states or partially disordered regions can be described as ensembles of rapidly interconverting conformations, some of which may be populated with greater preference. Structural study of these disordered states is important since little is known about the role of preferential conformations in the initiation of folding or the events that precede formation of extensive secondary and tertiary structure. Because kinetic folding intermediates exist only transiently and are not highly populated, a number of proteins have been studied under various degrees of denaturing conditions to create models for unfolded and intermediate states along the protein folding pathway (Neri et al., 1992; Logan et al., 1994; Alexandrescu et al., 1994a; Frank et al., 1995; Arcus et al., 1995; Buck et al., 1995). Experimental data indicate that, while some proteins show similar properties under all denaturing conditions, the unfolded states of many proteins vary markedly under different conditions (Dobson, 1992; Shortle, 1993). While denaturants are useful for enabling detailed studies of unfolded states, it is not clear what differences exist between the unfolded state under strong denaturing conditions and under folding conditions and whether studies of states generated under strong denaturing conditions have relevance for protein folding *in vivo*.

In the case of protein interactions, conformational changes often accompany binding. The role and extent of these conformational changes in biomolecular recognition are not well-established. It is generally believed that proteins exist in unique three-dimensional structures in order to perform specific biological functions. However, a protein may not adopt a folded structure even under native, folding conditions if stabilizing interactions are absent. There is accumulating evidence that a number of proteins exist in "native unfolded" states and that many of these are involved in protein−protein or protein−DNA interactions (Gast et al., 1995; Kriwacki et al., 1996; Schweers et al., 1994; Shen et al., 1996; Weinreb et al., 1996). Some of these proteins have been shown to exist in ordered states only when complexed with their targets; while the targets of other proteins have not yet been identified. Two examples of such proteins are the cyclin-dependent kinase (Cdk) inhibitor $p21^{Waf1/Cip1/Sdi1}$ (Kriwacki et al., 1996) and the transcriptional activation domain of the herpes virus protein VP16 (Shen et al., 1996). These proteins are transformed from unfolded, highly disordered states in their free forms into folded, more ordered states upon binding to their targets, Cdk2 and the basal transcription machinery, respectively. Another example is a 131-residue fragment of staphylococcal nuclease, Δ131Δ (Alexandrescu et al.,

† This work was supported through grants from the Medical Research Council of Canada (J.D.F.-K.). O.Z. acknowledges a graduate student fellowship from the Research Institute, Hospital for Sick Children, Toronto.

‡ Hospital for Sick Children and Department of Biochemistry, University of Toronto.

§ Department of Chemistry, University of Toronto.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1997.

S0006-2960(96)02762-6 CCC: $14.00

1994a), which is only partially folded under native conditions. In the presence of tight binding substrate analogues, however, it folds to a conformation similar to that of the wild-type molecule. These observations demonstrate that preferential binding of biological targets to the folded state can provide stabilization to the folded proteins and facilitate a disorder to order transition. Such binding-induced conformational changes from disordered states to ordered states may increase the specificity of protein−protein or protein−DNA recognition at the expense of binding affinity (Spolar & Record, 1994; Kriwacki et al., 1996).

Src homology 2 (SH2) and 3 (SH3[1]) domains are among a number of small, modular domains that are responsible for mediating specific classes of protein−protein interactions involved in regulating cell growth and differentiation [for a review, see Pawson (1995)]. SH2 domains are approximately 100 residues in length and bind to protein sequences containing a phosphotyrosine. SH3 domains are approximately 60 residues in length and mediate binding to specific proline-rich sequences. The *Drosophila* protein drk (Olivier et al., 1993; Simon et al., 1993) and its mammalian and *Caenorhabditis elegans* homologues, Grb2 and Sem-5, make up a family of small adapter proteins composed of a central SH2 domain flanked by two SH3 domains. The N-terminal SH3 domain specifically binds the guanine nucleotide exchange factor Sos *via* several repeats of the consensus sequence PXXP at the C terminus of Sos. The structures of a number of SH3 domains have been solved, either in free form or complexed with proline-rich peptides [for a review, see Kuriyan and Cowburn (1993)], including the N-terminal SH3 domain of Grb2 in complexed form (Wittekind et al., 1994; Terasawa et al., 1994; Goudreau et al., 1994) and in free form (Guruprasad et al., 1995).

The isolated N-terminal SH3 domain of drk exists in slow exchange on the NMR time scale between folded (F_{exch}) and unfolded (U_{exch}) states in aqueous buffer conditions close to physiological (Zhang et al., 1994; Zhang & Forman-Kay, 1995). A similar observation has also been reported for the isolated N-terminal SH3 domain of Grb2 (Goudreau et al., 1994). Earlier NMR experiments have shown that the F_{exch} state can bind to a polypeptide derived from the C-terminal region of Sos and that the U_{exch} state lacks long-range tertiary contacts but has some conformational preferences (Zhang & Forman-Kay, 1995). Addition of 2 M guanidine hydrochloride (Gdn) leads to a denatured state (U_{Gdn}) with a single set of resonances. Preliminary CD and fluorescence experiments suggest that this U_{Gdn} state is unchanged between 2 and 6 M Gdn (Y.-K. Mok, unpublished results). It is thus possible that the U_{Gdn} state represents the most extended state accessible in Gdn denaturant near neutral pH. This system provides an interesting model for studying the effects of binding of peptides or proteins on the equilibrium between folded and unfolded states and also for comparison of preferential conformations of unfolded states in near-physiological conditions with unfolded states under denaturing conditions.

NMR spectroscopy has become an important and powerful tool for detailed analysis of residual structure or structural preferences in unfolded or partially folded states (Wüthrich, 1994; Shortle, 1996). Because of the dynamic nature of these states, the significant dispersion of resonances observed for folded proteins collapses due to conformational averaging, especially for aliphatic ^1H and ^{13}C nuclei. Therefore, structural characterization of disordered states using conventional NMR methods has proven to be difficult. The realization that backbone ^{15}N and $^{13}C'$ (carbonyl) nuclei of unfolded proteins maintain reasonable dispersion and exhibit favorable relaxation properties has stimulated us to develop a suite of triple-resonance pulse sequences for structural characterization of unfolded or partially folded proteins (Zhang et al., 1997). The overall strategy of this approach is to transfer magnetization before or after an NOE mixing period to these relatively well-dispersed nuclei to resolve ambiguities which may arise from resonance overlap, enabling more unambiguous structural information to be obtained. Here we present results utilizing these pulse sequences as well as conventional NMR experiments on the two unfolded states of the drkN SH3 domain, the U_{exch} state in exchange with a folded state in low-salt aqueous buffer and the U_{Gdn} state in 2 M Gdn. Comparisons of structural features of these two unfolded states are presented, and implications of these results for protein folding are discussed.

MATERIALS AND METHODS

Sample Preparation. A proline-rich peptide with the sequence YRAVPPPLPPRR, derived from the *Drosophila* guanine nucleotide exchange factor Sos (residues 1139−1150), was synthesized according to standard procedures and was purified using a C18 semipreparative reverse-phase column on an HPLC system, in 0.1% v/v trifluoroacetic acid and a water/acetonitrile gradient. The peptide sample for titration was >95% pure by analytical reverse-phase HPLC, and the molecular mass of the peptide was confirmed by mass spectrometry. Peptide and protein concentrations were determined by quantitative amino acid analysis. The expression and purification of ^{15}N-labeled and ^{15}N,^{13}C-labeled drkN SH3 domain were as described previously (Zhang & Forman-Kay, 1995; Zhang et al., 1997). The protein sample for peptide titrations was a uniformly ^{15}N,10% ^{13}C-labeled protein sample initially prepared for the purpose of stereoassignment of prochiral methyl groups of Val and Leu residues (Neri et al., 1989). Variable-temperature experiments were done on a ^{15}N-labeled, 0.7 mM sample at pH 6.0 in 50 mM sodium phosphate and 90% H_2O/10% D_2O. NOESY experiments were performed on samples of either 1.5 mM ^{15}N,^{13}C-labeled or 0.7 mM ^{15}N-labeled protein, both at pH 6.0, in 90% H_2O/10% D_2O and 50 mM sodium phosphate at 5 °C. Later both the ^{15}N,^{13}C-labeled and ^{15}N-labeled samples were exchanged into buffers containing 2 M Gdn, 50 mM sodium phosphate (pH 6.0), and 90% H_2O/10% D_2O.

NMR Spectroscopy. NMR experiments were recorded on Varian UNITY Plus 600 and 500 MHz spectrometers equipped with triple-resonance pulsed field gradient probes

[1] Abbreviations: CD, circular dichroism; drk, downstream of receptor kinases, *Drosophila* homologue of mammalian Grb2; drkN SH3, N-terminal SH3 domain of drk; DSS, 2,3-dimethyl-2-silapentane-5-sulfonic acid; F_{exch}, folded state of the drkN SH3 domain existing in equilibrium with the unfolded state in low-ionic strength aqueous buffer; F_S, folded state of the drkN SH3 domain stabilized in 0.4 M sodium sulfate; Gdn, guanidine hydrochloride; HSQC, heteronuclear single-quantum coherence; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; SH3, src homology 3; 2D, two-dimensional; 3D, three-dimensional; TFE, 2,2,2-trifluoroethanol; U_{exch}, unfolded state of the drkN SH3 domain existing in equilibrium with the folded state in low-ionic strength aqueous buffer; U_{Gdn}, unfolded state of the drkN SH3 domain in 2 M Gdn.

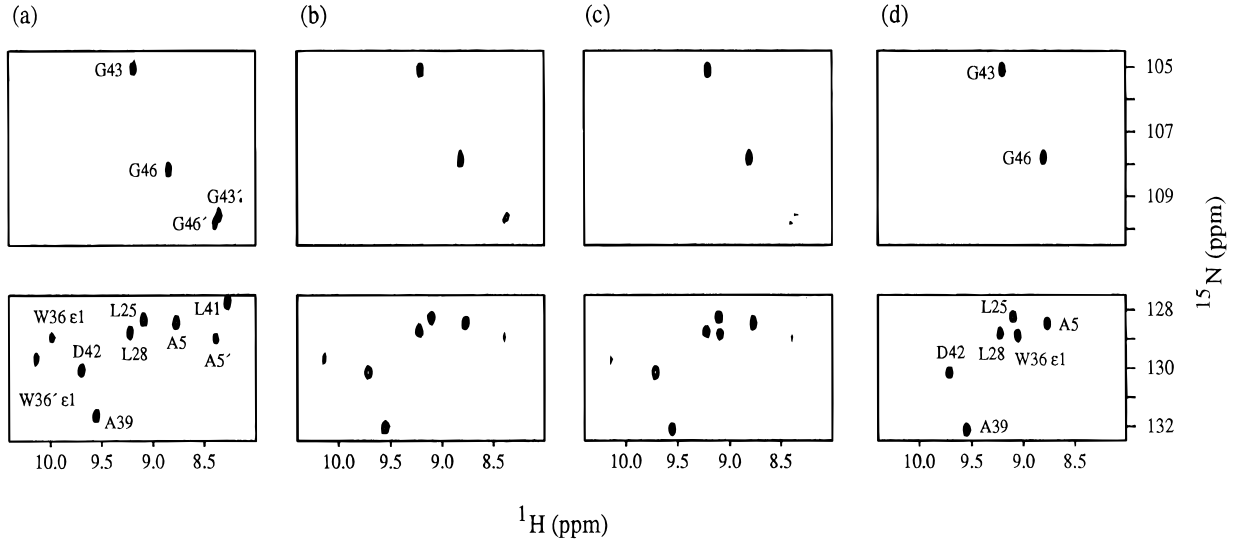

FIGURE 1: 2D ^{1}H−^{15}N HSQC spectra of the drkN SH3 domain with increasing amounts of the Sos-derived peptide. Molar ratios of peptide: protein are 0:1 (a), 0.5:1 (b), 1:1 (c), and 3:1 (d). Peaks labeled with prime signs correspond to residues from the unfolded state.

with actively shielded *z*-gradients and gradient amplifier units. ^{15}N−^{1}H HSQC spectra were taken with 64 and 512 complex points in t_1 and t_2, respectively, using a gradient sensitivity-enhanced approach (Kay et al., 1992; Schleucher et al., 1993). The (H)C(CO)NH−TOCSY and H(CCO)NH−TOCSY experiments (Logan et al., 1993; Grzesiek et al., 1993; Montelione et al., 1992) utilized 64, 32, and 576 complex points in t_1, t_2, and t_3, respectively, and were acquired with spectral widths of 9178.5 Hz for ^{13}C or 3600 Hz for ^{1}H, 1335.4 and 9000.9 Hz (F_1, F_2, and F_3) at 600 MHz, respectively. The HNHA experiment (Vuister & Bax, 1993) was recorded with spectral widths of 4000.0, 1335.4, and 9000.9 Hz (F_1, F_2, and F_3) at 600 MHz, utilizing 64, 48, and 576 complex points in t_1, t_2, and t_3, respectively. The N_{i+1}−NOESY−N_jH_j, (HN)CO_i−NOESY−N_jH_j, C_i−NOESY−TOCSY−$N_{j+1}H_{j+1}$, and ^{15}N HSQC−NOESY−HSQC experiments were recorded (Zhang et al., 1997), in addition to the conventional ^{15}N-edited NOESY−HSQC experiment. Experimental parameters for NOESY experiments and data processing procedures were similar to those reported previously (Zhang et al., 1997), and experiments were recorded with NOE mixing times of 250 ms. This long mixing time was used to maximize the intensity of NOE peaks derived from preferentially structured substates which may be present at a very low concentration. DSS was used as an internal reference at 0.00 ppm for proton chemical shifts. The ^{13}C chemical shifts are relative to external sodium acetate at 25.85 ppm. The ^{15}N chemical shifts are relative to external urea at 78.98 ppm.

Stabilization of the Folded State by Ligand Binding. For an equilibrium between two forms of a protein, a folded state (F) and an unfolded state (U),

$$\mathrm{U} \rightleftharpoons \mathrm{F} \tag{1}$$

the equilibrium constant (K_f) can be written as

$$K_f = [\mathrm{F}]/[\mathrm{U}]$$

Addition of a ligand, L, leads to formation of the complex FL. If it is assumed that the ligand only binds to the folded state at a single binding site with a dissociation constant K_d, then the equations

$$\mathrm{FL} \rightleftharpoons \mathrm{F} + \mathrm{L} \tag{2}$$

and

$$K_d = [\mathrm{F}][\mathrm{L}]/[\mathrm{FL}]$$

hold. Derivation of an expression to describe the effect of the ligand-binding equilibrium on the folding equilibrium yields

$$[\mathrm{FL}] = \frac{a + b + \left[\frac{K_d(1+K_f)}{K_f}\right] - \left[\left[a + b + \frac{K_d(1+K_f)}{K_f}\right]^2 - 4ab\right]^{1/2}}{2} \tag{3}$$

where a is the total concentration of protein and b is the total amount of ligand. This expression and a similar one for the amount of unfolded state were used to evaluate the effects of different dissociation constants on the folding equilibrium between the F_{exch} state and U_{exch} unfolded states of the drkN SH3 domain.

RESULTS

Peptide Titrations. HSQC spectra in Figure 1 illustrate titrations of a 12-residue proline-rich peptide derived from *Drosophila* Sos with a 0.5 mM, uniformly ^{15}N,10% ^{13}C-labeled sample of the drkN SH3 domain in 50 mM Tris at pH 7.5 and 18 °C. On the basis of structures of the N-terminal SH3 domain of Grb2 complexed with proline-rich peptides (Wittekind et al., 1994; Terasawa et al., 1994; Goudreau et al., 1994) and binding measurements for similar systems (Wittekind et al., 1994), it is expected that the folded state of the drkN SH3 domain will bind the proline-rich peptide in the minus orientation with a K_d of less than 50 μM. Panel a shows two portions of the HSQC spectrum of the free drkN SH3 domain with resonances for both the F_{exch} and U_{exch} (denoted with prime signs) states of the backbone amides of Gly 43 and Gly 46 (top part) and Ala 5 and the indole of Trp 36 (bottom part) labeled in the figure. Under

these conditions, the percentages of the F_{exch} and U_{exch} states are approximately 65 and 35%, respectively. Upon addition of the peptide, the intensities of resonances from the unfolded state decrease as seen in panel b. It should be noted that, while the intensity of the peak from the unfolded state of the Trp 36 indole decreases, the peak from the folded state of the Trp 36 indole broadens beyond detection due to intermediate exchange between the free and peptide-complexed folded states of the SH3 domain. Because of this exchange broadening, the intensities of many other resonances of the folded state (free plus complexed form) do not increase through the early points of the titration even though the equilibrium between unfolded and folded states is shifted toward the folded state of the drkN SH3 domain upon peptide binding.

When the stoichiometric ratio of peptide to protein approaches 1:1 (0.5 mM peptide and 0.5 mM protein), the resonances of the folded, complexed state begin to sharpen and increase in intensity and there is still a small fraction of unfolded state present as shown in panel c of Figure 1. These data are consistent with the model expressed in eq 3 (see Materials and Methods) and a K_d of 50 μM, predicting an equilibrium shifted largely to the folded, complexed form with roughly 0.05 mM unfolded protein remaining (intensities of the unfolded state decreased by a factor of 3.5). We have made no attempt to fit intensity or volume measurements to eq 3 in a quantitative manner. Instead, we have used eq 3 only as a semiquantitative guide due to difficulties in accurate measurements of the protein and peptide concentration and lack of a precise value for the K_d. The Trp 36 indole of the folded state broadens and disappears upon addition of peptide, and a new peak at 9.10 ppm in ^{1}H and 128.8 ppm in ^{15}N appears which we can assign to the folded, peptide-complexed form on the basis of spectral comparison. At a molar ratio for peptide to protein of 3:1, the equilibrium shifts almost fully to the folded, complexed form with more intense, sharp resonances and no observable resonances from the unfolded state. On the basis of our assumption of a 50 μM K_d, however, the unfolded form may still be present at a 0.01 mM concentration. This peptide titration experiment clearly shows that the binding of the biological target can stabilize the drkN SH3 domain.

Folded State Structure. The equilibrium between folded and unfolded states can also be shifted to a fully folded state by adding inorganic salts as demonstrated earlier (Zhang & Forman-Kay, 1995). The protein exists in a fully folded conformation (F_S state) in 400 mM Na_2SO_4. The structure of the F_{exch} state is essentially the same as that of the F_S state, on the basis of similarities in NOE patterns and chemical shifts (Zhang & Forman-Kay, 1995). Figure 2 shows a ribbon diagram of the F_S state of this SH3 domain at 18 °C determined by multidimensional NMR methods (unpublished results). The backbone consists of seven antiparallel β-strands (1−7), three variable loops, and two β-turns. Strand β2 is irregular as judged from the chemical shift index (see below) and forms a twisted β-hairpin with β3. The determined structure of the drkN SH3 domain is very similar to that of other SH3 domains, especially the N-terminal SH3 domain of Grb2 (Terasawa et al., 1994; Wittekind et al., 1994; Goudreau et al., 1994; Guruprasad et al., 1995) and provides a foundation for comparison of structural data from unfolded states of this domain.

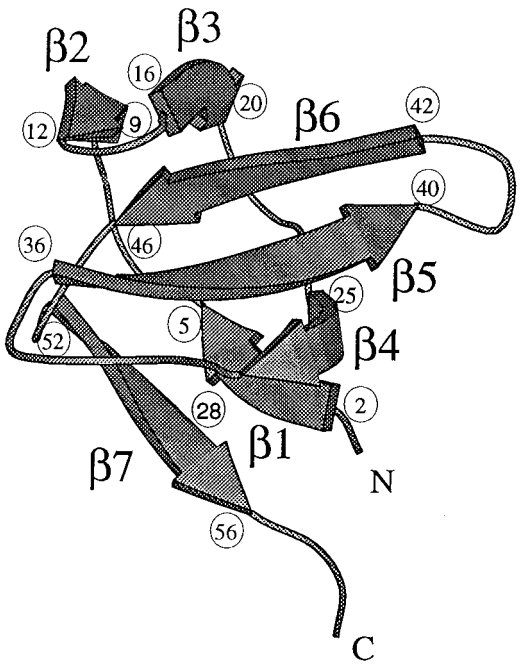

FIGURE 2: Backbone ribbon diagram (Kraulis, 1991) of the topology of the drkN SH3 domain based on preliminary structure calculations.

Chemical Shifts. Complete backbone and almost complete aliphatic side chain ^{1}H, ^{15}N, and ^{13}C assignments were determined for the F_{exch} folded state, as well as backbone and more limited side chain assignments for both the U_{exch} and U_{Gdn} unfolded states, using triple-resonance NMR experiments (Supporting Information). While in the process of resonance assignment of the U_{exch} state at 5 °C, we found that a number of resonances were significantly broadened, preventing completion of the assignments. In contrast, in the previous studies (Zhang & Forman-Kay, 1995) at 30 °C, no significant line broadening was observed for residues in the U_{exch} unfolded state. HSQC experiments at different temperatures were performed on a 0.7 mM sample at pH 6.0 with 50 mM sodium phosphate to investigate the temperature dependence of the broadening. Small portions of the HSQC spectra at various temperatures are shown in Figure 3. Relevant peaks are labeled by residue name and number, with resonances of the unfolded state denoted with prime signs. In this region of the spectra, both peaks from the folded and unfolded states of Ala 5, Ala 11, Leu 25, and Leu 28 are present, along with peaks from the folded states of additional residues. At higher temperatures (35 and 26 °C), resonances of neither the folded nor unfolded states experience significant line broadening. When the temperature is lowered, the line widths of Leu 25′ and Leu 28′ of the unfolded state become significantly broadened, clearly seen in the spectra at 14 and 6 °C, while the line widths and intensities of the resonances of the folded state are essentially unchanged. The Leu 25′ resonance disappears; it is not overlapped with Ala 39′ as revealed from 3D experiments which show only resonances of Ala 39′ at this chemical shift (unpublished observation). Overall, residues from Gln 23′ to Leu 28′ in the U_{exch} state experience severe line broadening. Similar variable-temperature experiments were also performed for the U_{Gdn} state, and no significant line broadening was observed (data not shown).

The available backbone $^1H^\alpha$ and $^{13}C^\alpha$ chemical shifts of both folded and unfolded states enable comparisons of secondary shifts, i.e. deviations of the observed chemical

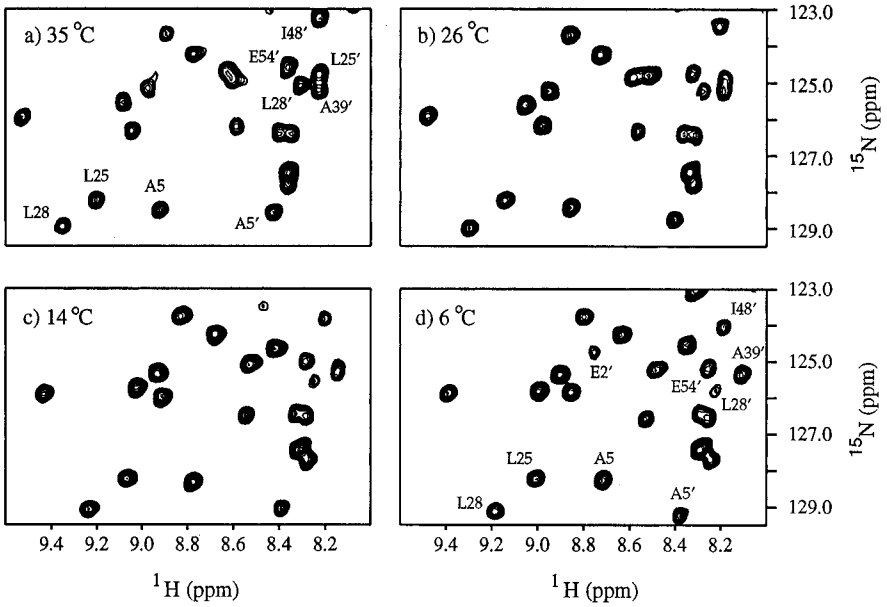

FIGURE 3: Portions of 2D ^{1}H−^{15}N HSQC spectra recorded at different temperatures to illustrate the resonance line broadening for residues Leu 25 and Leu 28 in the U_{exch} state. Selected resonances are highlighted. Peaks labeled with prime signs correspond to residues from the unfolded state. Note that E2′ is only visible at 6 °C due to solvent exchange broadening at higher temperatures.

shifts from random-coil values. Such secondary shifts can be used to define regular secondary structures in folded proteins and provide information on preferential conformations in unfolded proteins. Figure 4 illustrates the chemical shift differences of the backbone $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ resonances from random-coil values (Wishart et al., 1995) with corrections for proline nearest neighbor effects. In addition, in the case of $^{1}H^{\alpha}$ protons, these random-coil values have also been corrected for the aromatic contribution from neighboring residues at the $i - 2$, $i - 1$, $i + 1$, and $i + 2$ positions as suggested by Merutka et al. (1995) and Viguera et al. (1996).

The secondary shifts for the F_{exch} state (Figure 4a,b) readily define six strands (β1 and β3−β7) in the folded structure, while identification of an irregular β2 strand (from residue Phe 9 to Thr 12) requires complementary information from NOEs and coupling constants, although $^{13}C^{\alpha}$ of Ser 10 is significantly upfield of the random-coil value. The loop or turn regions connecting these strands are clearly shown by the upfield $^{1}H^{\alpha}$ and downfield $^{13}C^{\alpha}$ shifts for these residues relative to their random-coil values. The extent of secondary shifts in the F_{exch} state is very similar to that for $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ resonances of the N-terminal SH3 domain of Grb2 (M. Wittekind, personal communication) and for $^{1}H^{\alpha}$ resonances of the α-spectrin SH3 domain (Viguera et al., 1996).

The overall tendency of $^{1}H^{\alpha}$ shifts to lie upfield of random-coil values in the U_{exch} state is somewhat surprising (Figure 4c, note that data are missing for the region which is broadened). The upfield deviation is generally higher than that observed for peptide fragments derived from β-hairpin regions of the α-spectrin SH3 domain in aqueous buffer but is closer to that seen when 30% TFE is added to peptide fragments (Viguera et al., 1996). The $^{13}C^{\alpha}$ secondary shifts (Figure 4d) provide further evidence that the region from Asp 14 to Thr 22 substantially samples the α-region of conformational space, although Glu 16−Arg 20 are in a regular β-strand in the folded structure. Thus, these results demonstrate the formation of turn-like structures in rapid equilibrium with extended structure, which was also seen from NOE and coupling constant data (see below). Similar trends are observed for the U_{Gdn} state, although with somewhat less deviation from random-coil values, particularly between Asp 14 and Thr 22 (Figure 4e−h).

NOE Measurements. In order to characterize and compare the two unfolded states in more detail, newly developed NMR pulse sequences (Zhang et al., 1997) as well as conventional experiments have been utilized. In particular, we focused on the NOE connectivity patterns involving backbone NH protons. The N_{i+1}−NOESY−N_jH_j experiment can be used to identify NOEs between an α-proton of residue i and NH of residue j by providing correlations at ($\omega_{N_{i+1}}$, ω_{N_j}, ω_{NH_j}). Note that the ^{15}N chemical shift of residue $i + 1$ is recorded in F_1 in order to (a) avoid ambiguities from degeneracy of $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ shifts and (b) exploit the dispersion of backbone ^{15}N resonances. A detailed description of this pulse sequence is provided elsewhere (Zhang et al., 1996a). Figure 5a shows the application of this N_{i+1}−NOESY−N_jH_j experiment to the U_{exch} state of the drkN SH3 domain. A strip plot at the ^{15}N (F_2) and NH (F_3) chemical shifts of Ser 10 shows the sequential NOE between the H^{α} of Phe 9 and the NH of Ser 10 observed at the ^{15}N chemical shift of Ser 10 in F_1. The intraresidue H^{α}−NH NOE of Ser 10 is observed at the ^{15}N chemical shift (F_1) of Ala 11. A third NOE between the H^{α} of Asp 8 and Ser 10 is identified at the ^{15}N chemical shift of Phe 9. For comparison, a strip plot at the ^{15}N (F_2) and NH (F_3) chemical shifts of Ser 10 of the 3D ^{15}N-edited NOESY experiment is shown in Figure 5b. The almost identical H^{α} chemical shifts of Asp 8 and Phe 9 make unambiguous assignment of these NOEs impossible from this data set. In the case of several resonances having similar ^{15}N chemical shifts, a slightly modified experiment, the (HN)CO_i−NOESY−N_jH_j experiment which records the $^{13}C'$ chemical shift of residue i rather than the ^{15}N of residue $i + 1$, can be used to resolve ambiguities (Zhang et al., 1997).

The standard NMR experiment for assigning NOEs between amide protons in folded proteins is a 3D ^{15}N-edited NOESY experiment. The ^{15}N HSQC−NOESY−HSQC

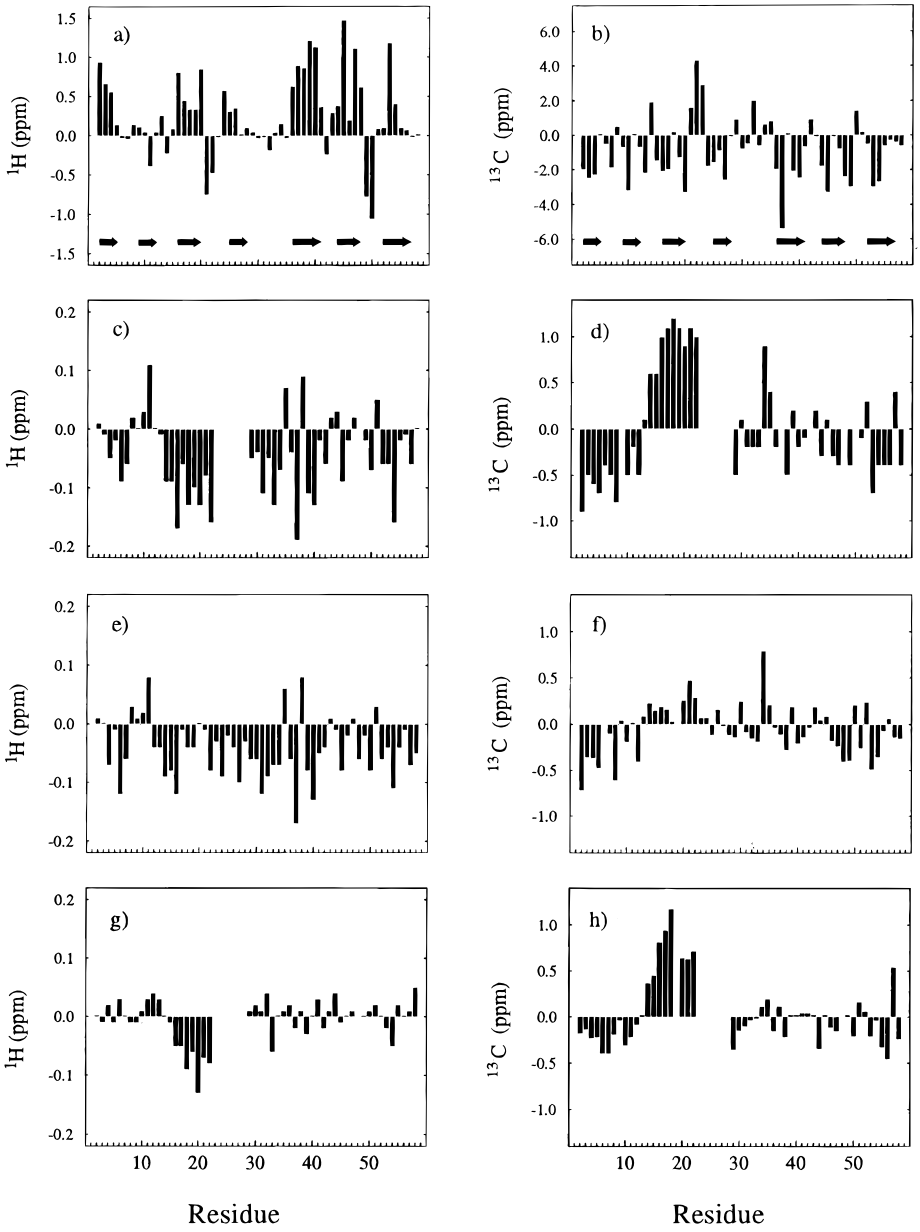

FIGURE 4: Comparison of chemical shifts of backbone H^α and C^α resonances of the F_{exch}, U_{exch}, and U_{Gdn} states to the random-coil (rc) values: (a and b) $F_{exch}-rc$, (c and d) $U_{exch}-rc$, (e and f) $U_{Gdn}-rc$, and (g and h) $U_{exch}-U_{Gdn}$. The values for both terminal residues are not shown since they are associated with end group terminal charge effects. The absence of data is due to the uncertainty in assignments of resonances either in line-broadened regions or a result of spectral overlap. (The rc H^α values are corrected for aromatic contributions as described in the text. Arrows in a and b correspond to the β-strands identified in the structure of the folded state.)

experiment is preferred for disordered states since it utilizes ^{15}N rather than ^{1}HN chemical shifts and exploits the increased dispersion of ^{15}N relative to ^{1}HN in unfolded proteins. This experiment was originally proposed to identify NOEs between amide protons in α-helical regions of proteins, which generally have poor NH dispersion (Ikura et al., 1990; Frenkiel et al., 1990). The observation of symmetry-related NOE cross-peaks at (ω_{N_i}, ω_{N_j}, ω_{NH_j}) and (ω_{N_j}, ω_{N_i}, ω_{NH_i}) is very helpful for resolving ambiguities arising from the possible degeneracy of ^{15}N chemical shifts in unfolded proteins. Figure 6 illustrates the resolving power of this experiment for the identification of NH−NH NOEs for the U_{exch} state in the drkN SH3 domain. Strips from the ^{15}N and NH chemical shifts of Thr 12, Glu 16, and Ser 18 are shown in Figure 6a and of Ala 39, Leu 41, and Gly 43 in Figure 6b. Strong sequential NOEs are clearly present for these residues. Particularly noteworthy are the medium-range NH−NH NOEs, including Thr 12−Glu 16 and Glu 16−Ser 18 NOEs (Figure 6a) and Ala 39−Leu 41 and Leu 41−Gly 43 NOEs (Figure 6b). The symmetric nature of the NOEs observed in this experiment allows confirmation of these assignments. While the ^{15}N-edited NOESY has intrinsically higher sensitivity, only a small number of unambiguous assignments can be made from this data due to very limited dispersion in NH resonances.

Figure 7 summarizes all the assigned unambiguous NOE data for the U_{exch} state (Figure 7a) and the U_{Gdn} states (Figure 7b). For both the U_{exch} and U_{Gdn} states at 5 °C, sequential $d_{\alpha N}(i,i+1)$ NOEs are observed throughout the protein sequence (except for the region from Gln 23 to Leu 28 in the U_{exch} due to line broadening) since $d_{\alpha N}(i,i+1)$ varies only between 2.2 and 3.6 Å for all possible conformations (Wüthrich, 1986). Thus, these NOE connectivities were not shown in the figure. Sequential $d_{NN}(i,i+1)$ NOEs are almost

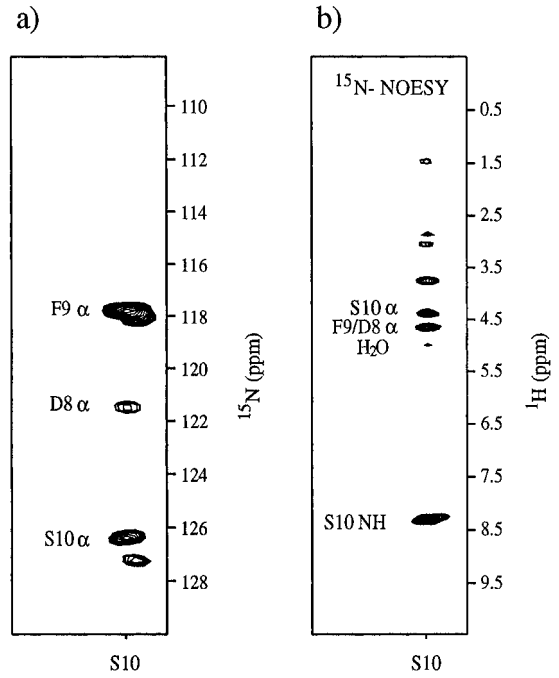

FIGURE 5: (a) Strip plot of the N_{i+1}−NOESY−N_jH_j spectrum of the U_{exch} state at the ^{15}N and NH chemical shifts of Ser 10 of the drkN SH3 domain. The three peaks represent NOEs from Ser 10 H^{α}, Phe 9 H^{α}, and Asp 8 H^{α} to Ser 10 NH. (b) Strip plot of the ^{15}N-edited NOESY spectrum at ^{15}N and NH chemical shifts of the same residue, demonstrating the overlap of the Asp 8 and Phe 9 H^{α} protons.

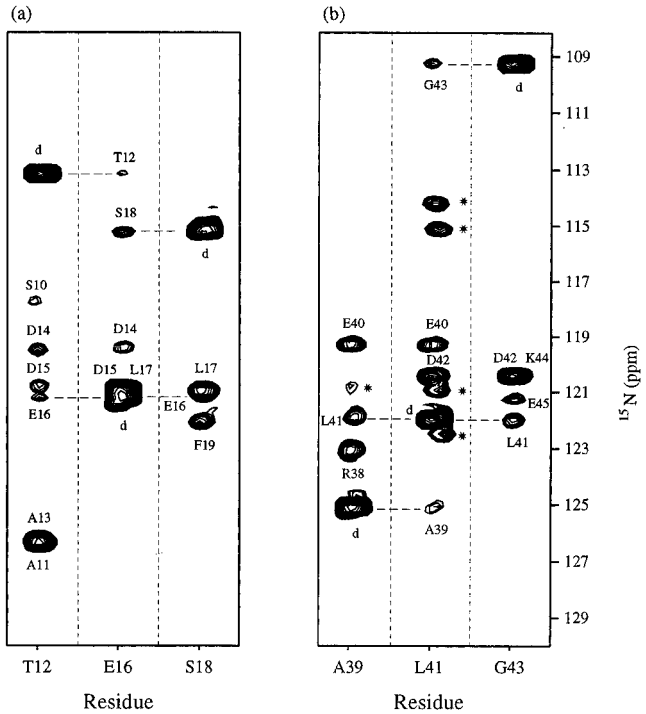

FIGURE 6: Strip plot from the 3D ^{15}N HSQC−NOESY−HSQC spectrum for (a) Thr 12, Glu 16, and Ser 18 and (b) Ala 39, Leu 41, and Gly 43 in the U_{exch} state of the drkN SH3 domain. "Auto" peaks are labeled with a d, and the NOE peaks are labeled with the site of origination of magnetization. The peaks labeled with asterisks are associated with residues having nearby ^{15}N and NH chemical shifts.

ubiquitous in both unfolded states. The sequential $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOEs themselves provide no information on secondary structure for short linear peptides as pointed out by Wright et al. (1988), and this holds for unfolded states of proteins as well. Nonetheless, the relative magnitude of these two types of sequential NOEs can give a qualitative measure of the relative population of backbone dihedral angles in the α- and β-regions of ϕ,φ space (Dyson & Wright, 1991). The R value [R = the intensity of $d_{NN}(i,i+1)$/ the intensity of $d_{\alpha N}(i,i+1)$] is indicative of the population of residues sampling the α-region (Waltho et al., 1993). For most of the drkN SH3 domain, the R value is 0.2−0.4, diagnostic of sampling both α- and β-regions of ϕ,φ space. In both the U_{exch} and U_{Gdn} unfolded states, the termini have the lowest R values, consistent with predominant sampling of extended conformations. There are three regions which have significantly larger R values in both unfolded states: Phe 9−Arg 20, Glu 31−Asn 35 and Leu 41−Lys 44. Many of these residues are in loop or turn regions in the folded state of the SH3 domain, but they also include residues from the irregular β2 strand and the regular β3 strand. $^3J_{HNH\alpha}$ coupling constants for the U_{exch} state at this temperature were measured by a quantitative J experiment, the HNHA (Vuister & Bax, 1993). Most of the $^3J_{HNH\alpha}$ coupling constants are approximately 6−7 Hz; however, Lys 6, Asp 14, Glu 16, Leu 17, Ser 18, Thr 22, Glu 31, Ser 34, Trp 36, Tyr 37, Glu 40, Ser 50, and Glu 54 have couplings below 6 Hz. A number of these residues having smaller ^3J coupling constants are in regions where there are turns or loops in the folded state, but others are from strands β3, β5, and β7. Gly and Ala residues have not been included in this analysis since in disordered regions their couplings are inherently lower due to steric effects (Wüthrich, 1986; Smith et al., 1996). The combination of secondary shifts, R values, and $^3J_{HNH\alpha}$ couplings for Glu 16−Arg 20 (which are in the β3 strand in the folded structure) provides strong evidence that these residues significantly sample the α-region of ϕ,φ space.

Comparison of the NOE patterns of the two unfolded states, U_{exch} and U_{Gdn}, shows that both the U_{exch} and U_{Gdn} states have ubiquitous $d_{\alpha N}(i,i+1)$ NOEs and significant numbers of intense $d_{NN}(i,i+1)$ NOEs, as mentioned above. A number of medium-range NOEs (predominantly $i,i+2$ and $i,i+3$) for both unfolded states have been unambiguously identified as described above. No NOEs beyond $i,i+4$ were unambiguously assigned. Three regions in the U_{exch} state, from His 7 to Ser 18, Met 30 to Arg 38, and Arg 38 to Ile 48, have significant numbers of $d_{\alpha N}(i,i+2)$ and $d_{NN}(i,i+2)$ connectivities as well as some longer-range NOEs. Note that a number of these residues are in strands β2, β3, β5, and β6 in the folded structure. In the U_{Gdn} state, fewer but still significant numbers of $i,i+2$ and even $i,i+3$ NOEs are present in regions close to where they are observed in the U_{exch} state, although the specific residues involved are sometimes different. The concentration of unfolded protein for the U_{exch} state is approximately 0.7 mM due to the roughly 1:1 ratio of F_{exch} and U_{exch} states while for the U_{Gdn} state is 1.5 mM, the total protein concentration. Thus, the larger number of observable intermediate-range NOEs in the U_{exch} state clearly demonstrates the greater degree of conformational preference in this state. However, the presence of these medium-range NOEs suggests that both states have preferred conformations of turn-like structures throughout the protein.

Although specific contacts between hydrophobic side chains in unfolded states may be present (Neri et al., 1992), the poor dispersion of aliphatic ^{1}H and ^{13}C chemical shifts makes unambiguous assignment of NOEs between aliphatic

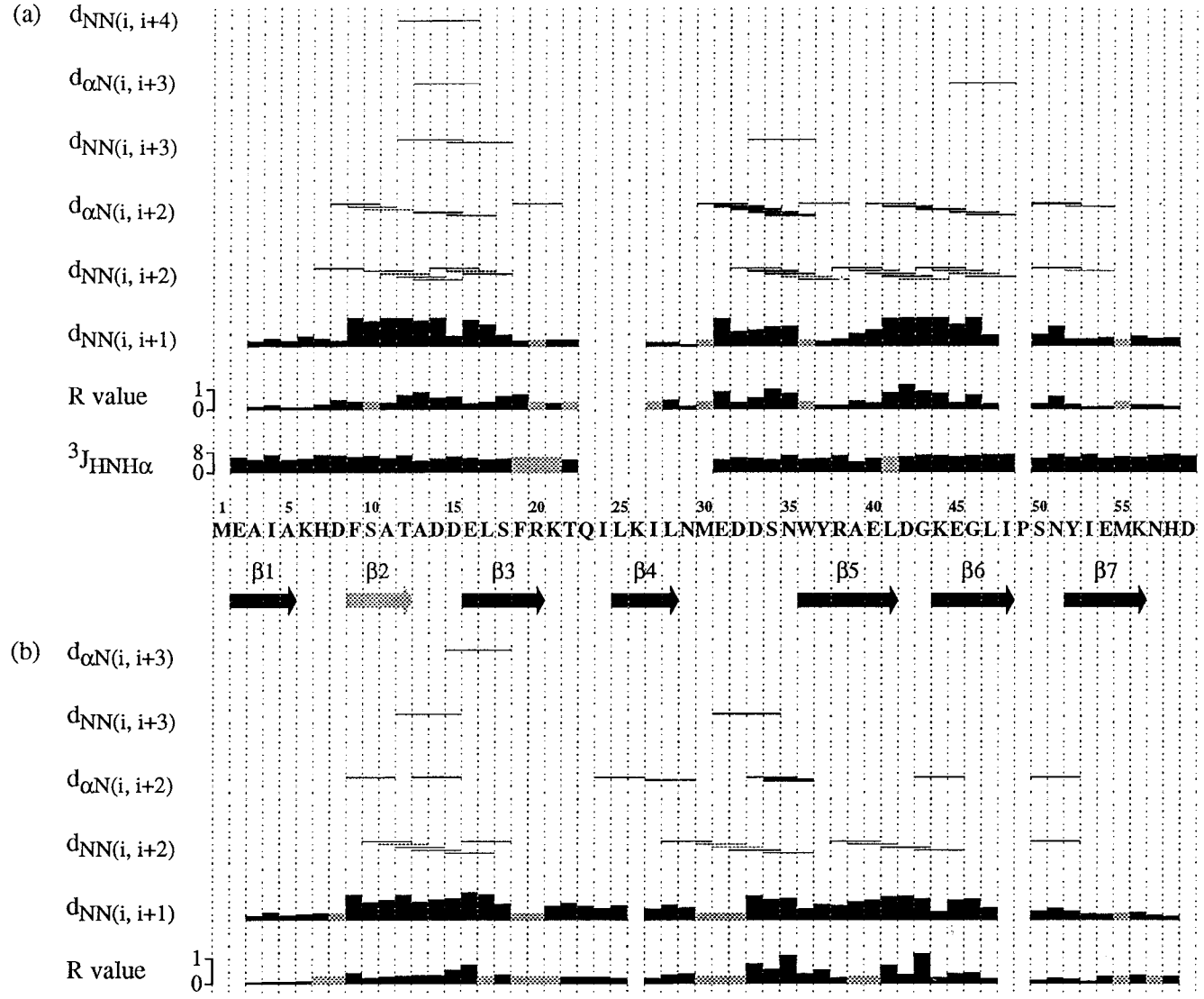

FIGURE 7: Summary of backbone NOE connectivities for the two unfolded states (a) U_{exch} and (b) U_{Gdn}. $^3J_{HNH\alpha}$ values for the U_{exch} state are also shown. R = the intensity of $d_{NN}(i,i+1)$/the intensity of $d_{\alpha N}(i,i+1)$. Arrows under the protein sequence refer to β-strands identified in the folded structure.

protons extremely difficult using conventional ^{13}C-edited NOESY spectra. New experiments with higher resolution were developed for assignment of NOEs between aliphatic protons (Zhang et al., 1996a). However, no medium- or long-range NOEs between aliphatic side chains could be identified for either the U_{exch} or U_{Gdn} unfolded states of the drkN SH3 domain. In particular, no $d\alpha\beta(i,i+3)$ NOEs, which are characteristic of stable helices, are observed. A few sequential aliphatic−aliphatic NOEs were identified. Note that application of these experiments to the partially folded deletion mutant Δ131Δ of staphylococcal nuclease did yield numerous sequential and medium-range aliphatic−aliphatic NOEs (Zhang et al., 1997, and unpublished results).

DISCUSSION

Peptide Binding. Results of the Sos peptide titration experiment confirm that the folded structure of the drkN SH3 domain is stabilized through binding to its biological target. Although the drkN SH3 domain is an isolated fragment from the protein drk, it is likely that there are minimal interactions between this SH3 domain and other domains in the intact protein on the basis of the crystal structure of the homologous protein Grb2 (Maignan et al., 1995). In addition, recent NMR chemical shift data for Grb2 complexed with target peptides for the SH2 and SH3 domains (Yuzawa et al., 1996) show no significant differences between the isolated SH2 or SH3 domain peptide complexes and these complexed domains in the intact protein. This suggests a lack of interactions between the domains. The binding site for biological targets on the drkN SH3 domain contains an aromatic patch with side chains of residues His 7, Phe 9, Trp 36, Pro 49, and Tyr 52 available for interaction with the proline-rich sequences. The surface of the SH3 domain contains a highly acidic patch, including the side chain of Glu 16, which is involved in electrostatic interaction with arginine(s) in the Sos peptide and thus in orienting the target (Feng et al., 1994). This patch may destabilize the domain by electrostatic repulsion. A study of the C-terminal SH3 domain of Sem-5, the *C. elegans* homologue of drk, also highlights the possible destabilization due to electrostatic repulsion of a cluster of negatively charged residues (Lim et al., 1994). These considerations argue that even under native conditions the N-terminal SH3 domain of intact drk could be destabilized in the absence of Sos or other binding targets.

Proteins exist in equilibrium between their folded and unfolded states, with typical marginal stabilities of about 5−15 kcal/mol (Pace, 1975), giving ratios of folded to unfolded conformations from 4.4×10^3 to 8.2×10^{10} at room temperature. This equilibrium can be changed dramatically, however, with different temperature and solvent conditions. For a trivial example, thermal denaturation of a protein gives an equilibrium constant at the transition temperature of 1, representing equal amounts of folded and unfolded protein. However, there are proteins which are highly disordered under physiological conditions with ΔG_f values greater than 0. At least in some cases, binding of their biological targets stabilizes the folded conformations of these proteins.

If we assume that only folded protein is able to bind target peptide, the stabilization from the binding can be described as (Schellman, 1975; Pace & McGrath, 1980)

$$\Delta G_s = -RT \ln[1 + [P]/K_d] \quad (4)$$

where K_d is the dissociation constant between peptide and protein, [P] is the concentration of peptide, R is the gas constant, and T is the temperature. If a K_d of 50 μM is assumed, at a total concentration of 0.5 mM peptide, the stabilization from binding is 1.4 kcal/mol and will therefore change the ratio of folded to unfolded from ~2:1 to ~7:1. The percentage of unfolded state would be less than 15% of total protein at the concentration of 0.5 mM peptide, which is consistent with the experimental observations. This ΔG_s value is significant compared to the free energy difference between folded and unfolded states of a protein. Given the large number of proteins which function by binding to other proteins, nucleic acids, or ligands and the published observations of disordered regions of intact proteins under physiological conditions (Gast et al., 1995; Kriwacki et al., 1996; Schweers et al., 1994; Shen et al., 1996; Weinreb et al., 1996), it is possible that a sizable number of proteins *in vivo* require binding of target for stabilization to form ordered and folded structures. Furthermore, depending on the availability of their biological binding targets during the normal changes in concentration of proteins and other molecules, such as during signaling, there may be significant populations of disordered states of proteins present under native conditions *in vivo*. Whether preferred conformations in these disordered states have any role in protein binding or recognition is not known.

Line Broadening. At low temperatures, the resonances of the U_{exch} unfolded state undergo intermediate exchange broadening due to motion on time scales (microseconds to milliseconds) which are comparable to the difference between chemical shifts in various substates sampled in the unfolded state ensemble. At higher temperatures, these motions become faster and no line broadening is observed. Attenuation of resonances in $^{15}N-^{1}H$ HSQC spectra could arise due to several different mechanisms. One possibility is derived from intermediate exchange between amide and solvent protons, which has a minimum rate at low pH and low temperature and is slow for residues that are hydrogen-bonded or buried. A second factor is intermediate exchange between different global conformational states, such as between the free and peptide-complexed states of the folded SH3 domain. A third possibility, observed in this case, is local conformational exchange due to dynamic behavior of particular residues or groups. Of course, the change of temperature has effects on the equilibrium between the folded and unfolded states as well as the line widths of peaks in HSQC spectra for both states. However, we found that these effects are very small and do not contribute significantly to the extreme line broadening observed for residues Gln 23−Leu 28 in the U_{exch} state.

The implications of line broadening on resonances of folded and unfolded proteins differ significantly, since motion in folded states is generally limited to explorations of a single deep well in the potential energy surface for the protein while unfolded proteins comprise a large number of rapidly interconverting conformations with transitions between multiple wells with comparable energies (Onuchic et al., 1995). Thus, narrow line widths reflecting "single" conformations are observed for folded proteins, whereas narrow line widths due to fast exchange between multiple equienergetic states are often seen for unfolded proteins. Residues in unstructured termini or loop regions in a folded protein often display line broadening due to solvent exchange, since these amides are solvent accessible and unlikely to be hydrogen-bonded, or as a result of local conformational exchange, since various configurations may be equally unstable given the lack of stabilizing interactions. Structural characterization of these residues *via* NOEs or couplings is hindered by the decrease in resonance intensity, but since they are not critical for defining the topology of proteins, this does not pose a serious problem for studies of folded proteins. Line broadening in unfolded proteins is due to stabilization of one or more of the conformational substates, leading to higher energy barriers and slower exchange between them. This may well be in the intermediate exchange broadening regime, and therefore, residues involved in stabilizing medium- or long-range (tertiary) interactions are most likely to experience line broadening. Examples include the hydrophobic box residues in the molten globule of α-lactalbumin (Alexandrescu et al., 1994b) and the residues involved in the β-meander in the Δ131Δ mutant of staphylococcal nuclease (Wang & Shortle, 1995). Resonance broadening in these regions makes it difficult to characterize these particularly important residues using NMR approaches.

The line broadening observed at low temperatures for Gln 23−Leu 28 implies restriction of conformational sampling in the U_{exch} state due to stabilizing interactions which lower the energy of particular sets of conformations relative to unstabilized, equienergetic substates and slow the rate of exchange between these substates in the unfolded ensemble. These interactions are destabilized upon addition of Gdn since line broadening was not observed in the U_{Gdn} state. It is worth recalling that previous studies at 30 °C showed notable chemical shift differences between the two unfolded states in this region and an extremely low value (<4 Hz) of the $^{3}J_{HNH\alpha}$ coupling constant for Leu 28 in the U_{exch} state which is increased in the U_{Gdn} state (Zhang & Forman-Kay, 1995). These differences provide further support for the greater persistence of preferential structure in the U_{exch} state than in the U_{Gdn} state in this region. However, ^{15}N relaxation studies demonstrate an additional millisecond to microsecond motional component needed to account for the dynamic properties of residues Leu 25 and Leu 28 in the U_{Gdn} state (Farrow et al., 1997), suggesting that at least some of these interactions are still present in the U_{Gdn} state, although to a lesser extent.

The residues which are line-broadened in the U_{exch} state include those found in the entire β4 strand in the folded state, which has interstrand hydrogen bonds with both β1 and β5, as well as two residues (Gln 23 and Ile 24) before the β4 strand. It is interesting to note that Lys 21, Thr 22, and Gln 23 (KTQ) are in the analogous position in the sequence of the SH3 domain as the RGD motif in the N- and C-terminal SH3 domains of Grb2 and also in the C-terminal SH3 domain of drk as well as the LGD and KGD sequences in the SH3 domains of PI3 kinase and α-spectrin, respectively. These RGD or similar residues in SH3 domains whose structures have been determined are superimposable with root-mean-square deviations similar to that for the core of the proteins (Guruprasad et al., 1995). This region is solvent accessible and was postulated in some SH3 domains to be involved in cytoskeleton organization. Residues KTQ in the drkN SH3 domain are located at the apex of a loop connecting strands β3 and β4. Leu 28, which is in a stable turn in the U_{exch} state on the basis of data recorded at 30 °C (Zhang & Forman-Kay, 1995), is the last residue of the β4 strand in the F_{exch} state. It is interesting to note that the regions which show the most marked differences between the two unfolded states and between the U_{exch} and random-coil states are residues surrounding this KTQ (RGD) motif. While the function of this region is not clear, the presence of stable interactions for these residues may suggest a role for recognition of target proteins in the unfolded state.

Implication for Initiation of Protein Folding. One of the least understood phases of protein folding is initiation, i.e. how protein folding starts and where folding initiates along the primary sequence. A general picture of an unfolded state is multiple conformations existing in rapid dynamic equilibrium, sampling both the α- and β-regions of ϕ,φ space. Structural characterization of unfolded proteins by NMR can provide evidence for structural preferences within these ensembles on a per-residue basis. While characterization of this equilibrium-populated structure is not directly relevant to identification of initiation sites, since detailed characterization of true kinetic intermediates is often not possible due to their transient nature, such information has been sought in order to provide insight into potential initiation sites (Arcus et al., 1995; Logan et al., 1994; Neri et al., 1992). These studies are particularly interesting for small proteins, whose folding behaviors have often been shown to obey simple two-state kinetics (Fersht, 1995, and references cited within), since preferential structures in unfolded states may be more likely to be productive when no intermediates accumulate in the folding pathways. Preliminary CD and fluorescence studies (Y.-K. Mok, unpublished results) suggest that the unfolding of this drkN SH3 domain is close to two-state.

Turns have been previously postulated to be likely sites for initiation of protein folding (Lewis et al., 1971; Zimmerman & Scheraga, 1977). Theoretical studies using Monte Carlo methods for on-lattice folding trajectories also implicate β-turns in the initiation of folding for many folding topologies (Skolnick & Kolinski, 1989). Molecular dynamics simulations suggest a nanosecond time scale for the folding and unfolding of a pentapeptide forming reverse turns in water (Tobias et al., 1991). Simulations involving Monte Carlo sampling techniques indicate that the stability of β-turns may not depend heavily on amino acid sequence and that β-turns may direct folding pathways despite their lack of stability (Yang et al., 1996).

Studies on peptide fragments derived from proteins composed of various secondary structural motifs as models for initiation of protein folding have been pioneered by Dyson and Wright [for reviews, see Wright et al. (1988) and Dyson and Wright (1993)]. Results from their groups and other groups have shown that sequences forming turns in native proteins also can form reasonably stable turns as isolated peptides (Dyson et al., 1988; Shin et al., 1993; Yao et al., 1994; Blanco et al., 1994). In a systematic study with an all α-helical protein, myohemerythrin, and a largely β-sheet protein, plastocyanin, different propensities for formation of local elements of secondary structures were suggested to be required for initiation of folding of the two structural classes of proteins (Dyson et al., 1992a,b). In general, peptides derived from helical segments of proteins were found to have a greater tendency to form helix or nascent helix, while those from β-sheet are mostly unstructured (i.e. extended). These results are often cited as evidence for the "framework" (Kim & Baldwin, 1990; Ptitsyn, 1992) or "diffusion-collision" (Karplus & Weaver, 1976) models of protein folding, with the argument that regions of a protein which show some degree of independent secondary structure in the absence of long-range tertiary interactions could act as initiation sites for the folding reaction. However, the low intrinsic stability of the helical or extended structures of these peptides does not warrant their description as formation of stable secondary structures.

From studies with peptide fragments derived from a 64-residue protein, chymotrypsin inhibitor 2 (CI2), Fersht's group proposed a "nucleation−condensation" model in which the overall structure condenses around an element of structure, the nucleus, that itself consolidates during the condensation (Otzen et al., 1994; Fersht, 1995). This notion that secondary and tertiary interactions form concurrently in this small protein is supported by data from protein engineering (Itzhaki et al., 1995). Moreover, it is consistent with a simulation of the folding of a lattice-model protein which proceeds by a nucleation−growth mechanism (Abkevich et al., 1994). Further evidence to support this model comes from the observation of simple two-state kinetics of folding for several small proteins, including the α-spectrin SH3 domain (Fersht, 1995; Viguera et al., 1994).

Many studies have demonstrated that structural preferences within unfolded states generated under different denaturing conditions may be different (Shortle, 1993; Logan et al., 1994; Arcus et al., 1995). Therefore, the coexistence of folded and unfolded states of the drkN SH3 domain under folding conditions provides a unique opportunity to probe residual structure within the unfolded state under folding conditions which may be relevant to initiation of folding for this β-sheet protein. Since detailed information of folding properties of β-sheet structure is more sparse than that of α-helical structure, studies of this β-sheet protein are of additional interest. It also enables us to compare conformational properties of the unfolded states in aqueous solution with those of the unfolded states under denaturing conditions. Additionally, results for the drkN SH3 domain can be compared with studies of peptides derived from β-sheet proteins.

Although the real role in the folding pathway for preferential structure found in unfolded states can only be established by kinetic studies, the simple two-state behavior of these small proteins with no accumulating intermediates

permits us to speculate on their possible roles in the initiation of folding. Our results demonstrate that both unfolded states preferentially sample the α-region of ϕ,φ space. The presence of many nonsequential backbone NOEs and secondary shifts indicating sampling of the α-region of ϕ,φ space supports the idea that turn-like structure may play a role in protein folding initiation of this β-sheet protein. Most of the turn-like structures are not stable except the turn at Leu 28 found previously at 30 °C (Zhang & Forman-Kay, 1995), and they are not necessarily populated at the same time. Therefore, the NOEs observed do not define cooperative helical structure. These turn-like structures are populated for both the U_{exch} and U_{Gdn} unfolded states and may function to bury local hydrophobic surfaces, maximize all possible stabilizing interactions, and restrict the search of conformational space in the folding process. These data are also consistent with the notion that local interactions stabilizing turn-like structures throughout these two unfolded states may function to bring parts of a polypeptide chain which are distant in primary sequence into close spatial proximity and therefore initiate protein folding. The potential for multiple initiation sites is suggested by the presence of intermediate-range NOEs over a range of sequence positions.

Recently, peptides derived from β-hairpin regions in the α-spectrin SH3 domain have been studied in an attempt to identify potential folding initiation sites (Viguera et al., 1996). The results indicate that none of the peptides preferentially populates to a large extent any particular secondary structure conformation. Even though there could be local interactions favoring native-like secondary structure, tertiary interactions appear to play a major role in defining the native secondary structure in the folded state. Our conformational analysis expands the observations from the studies of peptide fragments; the U_{exch} unfolded state of the drkN SH3 domain significantly populates turn-like conformations which are in rapid equilibrium with extended structures. The extent of non-native structures in this unfolded state is higher than that in most of the peptide fragments and close to that in some fragments from the α-spectrin SH3 domain with 30% TFE, a helix enhancement agent. It is surprising that in the context of the complete sequence there is not more native-like secondary structure. In addition, there is no evidence for long-range interactions in this unfolded state. While it has been suggested that the local amino acid sequence, rather than medium- to long-range interactions in the folded protein, determines the turn conformations in the folded state (Dyson et al., 1988), it is clear that stabilization of β-sheet proteins requires long-range interactions between β-strands. The fact that these interactions are not observed in either unfolded state is consistent with β-sheet stabilization occurring in the later stages of protein folding.

Our data suggest the potential roles of turn-like structures in the initiation of protein folding and the possibility of formation of such turns as an early event in the folding of β-sheet proteins. Because these turn-like structures in the unfolded states are not limited to regions in which the turns exist in the folded structure, our findings expand the results from studies of peptides derived from turn regions of folded proteins. Our results are not consistent with a simple framework model of protein folding but provide indirect support for a nucleation−condensation model including possible multiple initiation sites.

Finally, although the overall NOE connectivity patterns for these two unfolded states (U_{exch} and U_{Gdn}) are very similar, there are distinctive differences in detail, especially for medium-range NOE connectivities. Overall, there is more "structure" in the U_{exch} state than in the U_{Gdn} state, including the region from Gln 23 to Leu 28 broadened in the U_{exch} state. Generally, unlike the case for the drkN SH3 domain, unfolded states of proteins are not highly populated under native-like conditions, and it is necessary to use denaturing agents in order to create stable unfolded states for structural studies. The overall similarity of NOE connectivity patterns between the U_{exch} and U_{Gdn} states suggests that studies of unfolded proteins under denaturing conditions can provide insights relevant to *in vivo* unfolded states. However, structural differences between these unfolded states do exist, and this may have implications for interpretation of results on unfolded states under denaturing conditions.

ACKNOWLEDGMENT

We thank Dr. Lewis Kay (University of Toronto) for guidance with the NMR spectroscopy and enthusiastic support of the project, Dr. Gerry Gish (Mount Sinai Hospital, Toronto) for synthesis and help with purification of the proline-rich peptide used in the study, and Dr. David Shortle (Johns Hopkins University) for useful discussions.

SUPPORTING INFORMATION AVAILABLE

Tables summarizing the assignments of backbone and almost complete aliphatic side chain ^{1}H, ^{15}N, and ^{13}C resonances for the F_{exch} folded state and backbone ^{1}H, ^{15}N, and ^{13}C resonances for both the U_{exch} and U_{Gdn} unfolded states (12 pages). Ordering information is given on any current masthead page.

REFERENCES

Abkevich, V. I., Gutin, A. M., & Shakhnovich, E. I. (1994) *Biochemistry 33*, 10026−10036.
Alexandrescu, A. T., Abeygunawardana, C., & Shortle, D. (1994a) *Biochemistry 33*, 1063−1072.
Alexandrescu, A. T., Ng, Y. L., & Dobson, C. M. (1994b) *J. Mol. Biol. 235*, 587−599.
Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M., & Fersht, A. R. (1995) *J. Mol. Biol. 254*, 305−321.
Berg, J. M., & Shi, Y. (1996) *Science 271*, 1081−1085.
Blanco, F. J., Rivas, G., & Serrano, L. (1994) *Nat. Struct. Biol. 1*, 584−590.
Buck, M., Schwalbe, H., & Dobson, C. M. (1995) *Biochemistry 34*, 13219−13232.
Creighton, T. E. (1992) in *Protein Folding* (Creighton, T. E., Ed.) W. H. Freeman and Company, New York.
Dobson, C. M. (1992) *Curr. Opin. Struct. Biol. 2*, 6−12.
Dyson, H. J., & Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−538.
Dyson, H. J., & Wright, P. E. (1993) *Curr. Opin. Struct. Biol. 3*, 60−65.
Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., & Wright, P. E. (1988) *J. Mol. Biol. 201*, 161−200.
Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992a) *J. Mol. Biol. 226*, 795−817.
Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., & Wright, P. E. (1992b) *J. Mol. Biol. 226*, 819−835.
Farrow, N. A., Zhang, O., Forman-Kay, J. D., & Kay, L. E. (1997) *Biochemistry* (in press).
Feng, S., Chen, J. K., Yu, H., Simon, J. A., & Schrieber, S. L. (1994) *Science 266*, 1241−1247.

Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869−10873.
Frank, M. K., Clore, G. M., & Gronenborn, A. M. (1995) *Protein Sci. 4*, 2605−2615.
Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., & Feeney, J. (1990) *J. Magn. Reson. 90*, 420−425.
Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Muller-Frohne, M., Zirwer, D., Czarnecki, J., & Damaschun, G. (1995) *Biochemistry 34*, 13211−13218.
Goudreau, N., Cornille, F., Duchesne, M., Parker, F., Tocque, B., Garbay, C., & Roques, B. P. (1994) *Nat. Struct. Biol. 1*, 898−907.
Grzesiek, S., Anglister, J., & Bax, A. (1993) *J. Magn. Reson. B 101*, 114−119.
Guruprasad, L., Dhanaraj, V., Timm, D., Blundell, T. L., Gout, I., & Waterfield, M. D. (1995) *J. Mol. Biol. 248*, 856−866.
Ikura, M., Bax, A., Clore, G. M., & Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 9020−9022.
Itzhaki, L. S., Otzen, D. E., & Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
Karplus, M., & Weaver, D. L. (1976) *Nature 260*, 404−406.
Kay, L. E., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., & Wright, P. E. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11504−11509.
Kuriyan, J., & Cowburn, D. (1993) *Curr. Opin. Struct. Biol. 3*, 828−837.
Lewis, P. N., Momany, F. A., & Scheraga, H. A. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2293−2297.
Lim, W., Fox, R. O., & Richards, F. M. (1994) *Protein Sci. 3*, 1261−1266.
Logan, T. M., Olejniczak, E. T., Xu, R. X., & Fesik, S. W. (1993) *J. Biomol. NMR 3*, 225−231.
Logan, T. M., Theriault, Y., & Fesik, S. W. (1994) *J. Mol. Biol. 236*, 637−648.
Maignan, S., Guillotesu, J.-P., Fromage, N., Arnoux, B., Becquart, J., & Ducruix, A. (1995) *Science 268*, 291−293.
Merutka, G., Dyson, H. J., & Wright, P. E. (1995) *J. Biomol. NMR 5*, 14−24.
Montelione, G. T., Lyons, B. A., Emerson, S. D., & Tashiro, M. (1992) *J. Am. Chem. Soc. 114*, 10974−10975.
Neri, D., Szyperski, T., Otting, G., Senn, H., & Wüthrich, K. (1989) *Biochemistry 28*, 7510−7516.
Neri, D., Billeter, M., Wider, G., & Wüthrich, K. (1992) *Science 257*, 1559−1563.
Olivier, J. P., Raabe, T., Henkemeyer, M., Dickerson, B., Mbamalu, G., Margolis, B., Schlessinger, J., Hafen, E., & Pawson, T. (1993) *Cell 73*, 179−191.
Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z., & Socci, N. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 3626−3630.
Otzen, D. E., Itzhaki, L. S., Elmasry, N. F., Jackson, S. E., & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10422−10425.
Pace, C. N. (1975) *Crit. Rev. Biochem. 3*, 1−43.
Pace, C. N., & McGrath, T. (1980) *J. Biol. Chem. 255*, 3862−3865.
Pawson, T. (1995) *Nature 373*, 573−580.
Ptitsyn, O. B. (1992) *Curr. Opin. Struct. Biol. 2*, 13−20.
Schellman, J. A. (1975) *Biopolymers 14*, 999−1018.
Schleucher, J., Sattler, M., & Griesinger, C. (1993) *Angew. Chem., Int. Ed. Engl. 32*, 1489−1491.
Schweers, O., Schonbrunn-Hanebeck, E., Marx, A., & Mandelkow, E. (1994) *J. Biol. Chem. 269*, 24290−24297.
Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., & Knutson, J. R. (1996) *J. Biol. Chem. 271*, 4827−4837.
Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E., & Dyson, H. J. (1993) *Biochemistry 32*, 6348−6355.
Shortle, D. (1993) *Curr. Opin. Struct. Biol. 3*, 66−74.
Shortle, D. (1996) *Curr. Opin. Struct. Biol. 6*, 24−30.
Simon, M. A., Dodson, G. S., & Rubin, G. M. (1993) *Cell 73*, 169−177.
Skolnick, J., & Kolinski, A. (1989) *Annu. Rev. Phys. Chem. 40*, 207−235.
Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., & Dobson, C. M. (1996) *J. Mol. Biol. 255*, 494−506.
Spolar, R. S., & Record, M. T., Jr. (1994) *Science 263*, 777−784.
Terasawa, H., Kohda, D., Hatanaka, H., Tsuchiya, S., Ogura, K., Nagata, K., Ishii, S., Mandiyan, V., Ullrich, A., Schlessinger, J., & Inagaki, F. (1994) *Nat. Struct. Biol. 1*, 891−897.
Tobias, D. J., Mertz, J. E., & Brooks, C. L., III (1991) *Biochemistry 30*, 6054−6058.
Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., & Serrano, L. (1994) *Biochemistry 33*, 2142−2150.
Viguera, A. R., Jimenez, M. A., Rico, M., & Serrano, L. (1996) *J. Mol. Biol. 255*, 507−521.
Vuister, G. W., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 6337−6347.
Wang, Y., & Shortle, D. (1995) *Biochemistry 34*, 15895−15905.
Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., & Lansbury, P. T., Jr. (1996) *Biochemistry 35*, 13709−13715.
Wilson, I. A., & Stanfield, R. L. (1994) *Curr. Opin. Struct. Biol. 4*, 857−867.
Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
Wittekind, M., Mapelli, C., Farmer, B. T., II, Suen, K.-L., Goldfarb, V., Tsao, J., Lavoie, T., Barbacid, M., Meyers, C. A., & Mueller, L. (1994) *Biochemistry 33*, 13531−13539.
Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167−7175.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Wüthrich, K. (1994) *Curr. Opin. Struct. Biol. 4*, 93−99.
Yang, A.-S., Hitz, B., & Honig, B. (1996) *J. Mol. Biol. 259*, 873−882.
Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E., & Dyson, H. J. (1994) *J. Mol. Biol. 243*, 736−753.
Yuzawa, S., Yokochi, M., Tsuchiya, S., Teresawa, H., Kohda, D., Schlessinger, J., Miura, K., & Inagaki, F. (1996) in *XVII*th*International Conference on Magnetic Resonance in Biological Systems*, Keystone, CO.
Zhang, O., & Forman-Kay, J. D. (1995) *Biochemistry 34*, 6784−6794.
Zhang, O., Kay, L. E., Olivier, J. P., & Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
Zhang, O., Forman-Kay, J. D., Shortle, D., & Kay, L. E. (1997) *J. Biomol. NMR* (in press).
Zimmerman, S. S., & Scheraga, H. A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4126−4129.

BI9627626