3246 *Biochemistry* **2010,** *49,* 3246–3253
DOI: 10.1021/bi902125j

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Residue-Specific Description of Non-Native Transient Structures in the Ensemble of Acid-Denatured Structures of the All-β Protein c-src SH3[†]

Heike I. Rösner and Flemming M. Poulsen*

Structure Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark

Received December 11, 2009; Revised Manuscript Received March 8, 2010

ABSTRACT: Secondary chemical shift analysis has been used to characterize the unfolded state of acid-denatured c-src SH3. Even though native c-src SH3 adopts an all-β fold, we found evidence of transient helicity in regions corresponding to native loops. In particular, residues 40−46, connecting the n-src loop to the third β-strand, exhibited an apparent helicity of nearly 45%. Furthermore, the RT loop and the diverging turn appeared to adopt non-native-like helical conformations. Interestingly, none of the residues found in transient helical conformations exhibited significant ϕ-values [Riddle, D. S., et al. (1999) *Nat. Struct. Biol. 6*, 1016−1024]. This indicated that the transient helicity has no influence or only a weak influence on the actual protein folding reaction. The residual structural propensities were compared to those of other SH3 domains, revealing heterogeneity in the unfolded ensemble that clearly contrasts with the conserved character of the topology of native state and transition state ensembles typical for SH3 domains.

The protein folding problem has occupied many scientists for several decades. It is widely accepted that studies of the ensemble of unfolded structures may assist in the characterization of the protein folding reaction (*1*). Detailed studies include the characterization of the unfolded ensembles of apomyoglobin, ACBP, Im7, or protein L (*2*−*5*). All four proteins consistently exhibit locally constrained regions within the unfolded ensembles. The transiently populated structures clearly resemble early folding intermediates or transition state ensembles. However, the current research is strongly biased toward unfolded ensembles of helical proteins or mixed α/β proteins. This might obscure the general portrayal of unfolded ensembles. Hence, we aimed to examine the ensemble of unfolded structures of an all-β protein at a residue-specific level. As a model system, we have chosen to analyze the secondary structural propensities of the unfolded state of c-src SH3.[1]

Src homology 3 domains (SH3) are small domains of ∼50−70 amino acids specialized in promoting protein−protein interactions. Even though they exhibit a very low degree of sequence homology (*6*), they reliably show the same fold topology, a sandwich structure of two small orthogonal three-strand β-sheets packed against each other (see the inset of Figure 1b). The β-strands are connected by four characteristic loops termed (a) the RT loop, (b) the diverging turn, (c) the n-src loop, and (d) the distal hairpin (*7*). Several variants, such as the spectrin, fyn, and c-src SH3 domains, have been subjected to extensive studies that aimed to characterize the folding pathways and their respective transition states (*8*−*10*). Despite their low degree of sequence homology of < 30−40%, these SH3 domains display a highly conserved transition state ensemble (TSE), which appears to be localized in the RT loop and the second half of the protein. c-src SH3 has the most localized TSE of this group, and residues with large ϕ-values such as E30, V35, A45, S47, L48, T50, T53, and I56 all reside in the second half of the primary sequence (*8*).

In contrast to the highly conserved topology, SH3 domains do exhibit a wide range of backbone dynamics and stability profiles. Slow local and global unfolding has been characterized for a large set of SH3 domains, which revealed diverse and sequence-independent partial unfolding reactions (*11*). The overall stability of the SH3 fold can also vary remarkably. The N-terminal SH3 domain of the *Drosophila* signal transduction protein drk (drkN SH3) domain provides an excellent model for the examination of the unfolded state under physiological conditions (*12*−*14*). Convincing evidence exists that drkN SH3 forms local native-like structure in the region corresponding to the last two β-strands. Interestingly, local non-native secondary structure propensity in the regions corresponding to the RT loop could also be detected (*13*).

The folding reaction of c-src SH3 was originally described as a classical two-step folding process (*8*). However, a series of recent studies shows that the folding pathway might be much more complex. First, the folding reaction of c-src SH3 was explored under physiological conditions using a reversible fragment assembly method. This theoretical approach revealed that the formation of a large amount of α-helix precedes the previously characterized c-src SH3 folding reaction (*15*). Subsequently, the folding kinetics of c-src SH3 at 4 °C and pH 3.0 were explored using CD spectroscopy. This study confirmed that a large amount of α-helix is populated at the very early onset of the folding (*16*). Interestingly, a recent study of hNck2 SH3 shows that this SH3 can be reversibly transformed into a stable and non-native helical state by acid unfolding (*17*). In addition, a single mutation in the c-src SH3 background, A45G, can form a stable

[†]This work was supported by The Danish Research Council for Natural Sciences Grant 272-08-0500, by Carlsberg Foundation Grant 2007_01_0268, and by The John and Birthe Meyer Foundation (F.M.P.).

*To whom correspondence should be addressed: Department of Biology, Ole Maaløesvej 5, University of Copenhagen, 2200 Copenhagen N, Denmark. Phone: +45 353 22077. Fax: +45 3532 2128. E-mail: fmp@bio.ku.dk.

[1]Abbreviations: CD, circular dichroism; SH3, Src homology domain 3; NMR, nuclear magnetic resonance; TSE, transition state ensemble; Ssp, secondary structure propensity score.

 Published on Web 03/10/2010

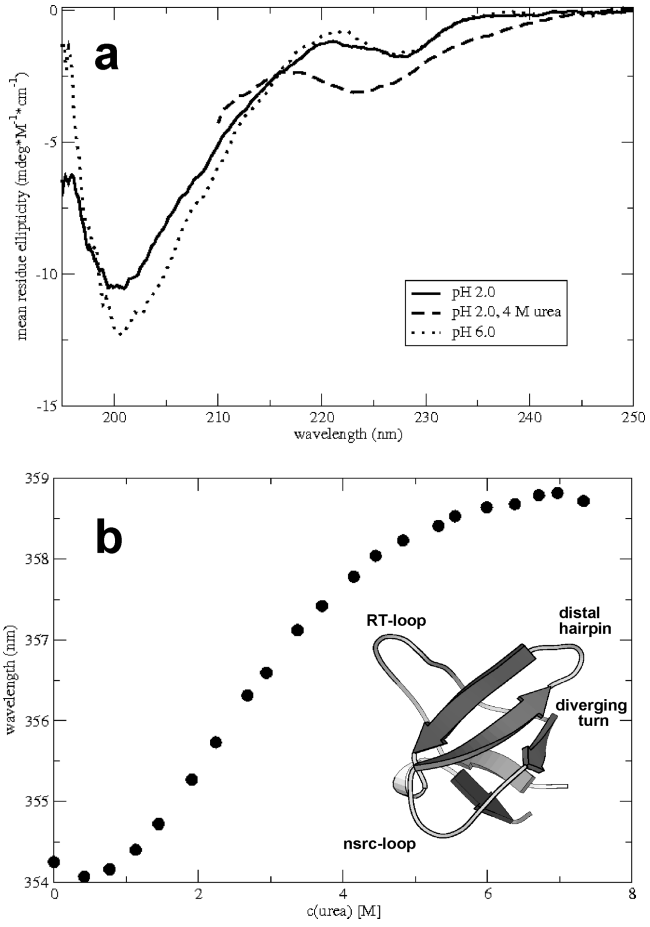

FIGURE 1: (a) CD measurements of folded c-src SH3 at 25 °C in 250 mM sodium phosphate (pH 2.0) in the absence of urea (—) and the presence of 4 M urea (– – –) and pH 6.0 (· · ·). (b) Wavelength at the center of mass of the internal fluorescence of c-src SH3 as a function of urea concentration in 250 mM sodium phosphate (pH 2.0). The protein concentration was 20 μM for both experiments. The inset shows a backbone ribbon diagram of c-src SH3 (Protein Data Bank entry 1SRL) using Molscript (*42*).

α-helix rich folding intermediate at pH 3.0 (*18*). High-resolution structures of these helical states are still missing.

Here, we examined the acid-denatured ensemble of c-src SH3 at the molecular level by NMR spectroscopy. The primary focus of our work was secondary chemical shift analysis. The choice of an ideal model for determining the chemical shifts of the random coil state is hereby crucial for the determination of very small deviations from true random coil behavior, but many available reference sets for random coil chemical shifts (*19*−*22*) are biased by the underlying experimental conditions. Hence, instead of focusing on such absolute reference sets, the analysis performed in this study uses an approach that compares several data sets recorded at increasing concentrations of denaturant. In this manner, we obtain a maximum denatured state from which the "random coil" chemical shifts can be used as an internal reference (*23*). This can monitor a relative increase in the level of formation of transient structure in the unfolded ensemble and facilitate the interpretation of secondary chemical shifts.

Using this method, a residue-specific description of the transient secondary structures in the ensemble of unfolded c-src SH3 was obtained. Regions of apparent transient structure were localized outside the regions of native secondary structure, and they did not correspond to regions previously identified as being involved in TSE formation. These results were compared to results obtained for other well-characterized SH3 domains. This revealed that the unfolded states clearly differ in their structural preferences despite the conserved nature of their native structures and their transition state ensembles.

MATERIALS AND METHODS

CD and Fluorescence Measurements. c-src SH3 was expressed and purified as described previously (*8* and references cited therein). The His tag was not removed for this study. CD measurements of c-src SH3 were conducted at a protein concentration of 20 μM at 25 °C in 250 mM sodium phosphate at (1) pH 2.0, (2) 4 M urea and pH 2.0, and (3) pH 7.0 using a JASCO J810 spectropolarimeter. The internal fluorescence of a 20 μM solution of c-src SH3 was measured as a function of urea concentration in 250 mM sodium phosphate (pH 2.0) using a Perkin-Elmer LS50B spectrometer. Generally, five spectra were recorded and averaged. The center of mass of the respective spectrum was used for plotting the urea dependency.

NMR Experiments. The $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts of ^{13}C- and ^{15}N-labeled c-src SH3 at a protein concentration of 0.2 mM were recorded in 250 mM sodium phosphate at pH 2.0 and 25 °C and seven different urea concentrations (0, 2, 3, 4, 6, 8, and 9.5 M), using standard ^{15}N-edited HSQC and ^{13}C,^{15}N-edited HNCA, HN(CO)CA, HNCO, and HN(CA)CO experiments on a Varian Inova 800 MHz spectrometer. The actual urea concentrations were confirmed by measurement of the refractive index of the solution. The proton chemical shifts were referenced internally to DSS, and the carbon and nitrogen chemical shifts were referenced indirectly.

Urea Coefficients. The chemical shifts for all four nuclei, $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N, were fitted as a linear function of urea concentration. The *y*-intercept was regarded as the chemical shift in the absence of denaturant. The slopes obtained from the fitting procedures described above were scaled and combined according to ref *27*, using the reported scaling factors for the combination of $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts:

$$\mathrm{coeff}_{\mathrm{urea}} = (0.82|\Delta\delta_{\mathrm{Ca}}| + 0.79|\Delta\delta_{\mathrm{Co}}| + 0.28|\Delta\delta_{\mathrm{NH}}| + 0.22|\Delta\delta_{^{15}\mathrm{N}}|)/4 \quad (1)$$

Determination of the Site-Specific Fraction of Helix/Turn Formation. The fraction of helix, F_i, at site *i* was determined as the ratio between the observed secondary chemical shift value using intrinsically referenced "random coil" shifts (δ_{irc}) as a reference ($\delta_{iobs} - \delta_{irc}$) and the average chemical shift value of the residue type when occurring in α helices in native proteins, as obtained from BioMagResBank (BMRB) and the site-specific intrinsically referenced random coil shifts (δ_{irc}).

$$F_i = (\delta_{iobs} - \delta_{irc})/[\delta_{\mathrm{RT}(\alpha)} - \delta_{irc}] \quad (2)$$

From this, the residue-specific free energy (ΔG) can be estimated by

$$\Delta G = -RT \ln K = -RT \ln[F_i/(1 - F_i)] \quad (3)$$

RESULTS

Equilibrium Denaturation Monitored by Internal Fluorescence and NMR Spectroscopy. As reported previously (*16*, *24*), the stability of c-src SH3 decreases and the midpoint upon GdmHCl denaturation shifts from 2.76 M at pH 6.0 to 2.1 M at pH 3.0, but the native state of c-src SH3 still remains remarkably stable even under very acidic conditions. Indeed, the

CD spectrum of c-src SH3 at pH 2.0 in Figure 1a shows that c-src SH3 remains almost fully folded despite the low pH value. To determine the optimal denaturing conditions for c-src SH3, a urea-induced denaturation of c-src SH3 was followed by internal fluorescence at pH 2.0. The plot in Figure 1b shows the center of mass as a function of the respective urea concentration. The centers of mass of the fluorescence spectra were 354.2 and 358.8 nm in the absence and presence of 7 M urea, respectively. The sigmoidal transition curve obtained for the stepwise unfolding under equilibrium conditions clearly indicates cooperative unfolding. However, a well-defined native baseline is missing.

Subsequently, the unfolded state of c-src SH3 was characterized by NMR spectroscopy to describe the profile of the denatured state on a residue-specific level. At pH <2.8, two full sets of peaks were present in the HSQC spectrum in the absence of urea, showing that the exchange between native state N and unfolded ensemble U is slow on the time scale of NMR experiments. In addition, the reported unfolding and folding rates of c-src SH3 at pH 2.0 ($\sim$7 and 1 s^{-1}, respectively) (*25*) confirm that the chemical exchange between the N and U states is comparatively slow on the time scale of NMR experiments. Hence, the influence of the chemical exchange between the folded and unfolded state on the appearance of the NMR spectra can be excluded. To gain better resolution of the peaks arising from the unfolded state, we measured the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts in a series of HN-detected three-dimensional NMR spectra at pH 2.0 and 25 °C for six different urea concentrations between 2 and 9.5 M. In the absence of urea, the amount of unfolded protein present in equilibrium at pH 2.0 was estimated to be 5−10% using the relative peak intensity. From 0 to 3 M urea (pH 2.0), peaks arising from denatured as well as native protein were present in the HSQC spectra. At urea concentrations of $\geq$4 M (pH 2.0), only peaks arising from fully denatured protein were detectable. The assigned chemical shifts have been deposited in the BioMagResBank as accession number 16750.

Determination of Urea Coefficients. Following the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts over this wide range of urea concentrations, several of the peaks changed their position significantly. These changes in chemical shift were not uniform for all residues. In addition, this behavior appeared to be independent of residue type, excluding a certain urea dependence enforced by the respective side chain. The inset of Figure 2 shows the $^{13}C^{\alpha}$ chemical shifts of Q52 as a function of urea concentration. Similarly, most of the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts exhibited a nearly linear dependence over the selected range of urea concentrations. A linear fit of the data yielded a slope that could be used as a measure for the urea dependence of the chemical shifts. The respective *y*-intercept yielded the extrapolated chemical shift in the absence of urea. The slopes obtained from the plots of chemical shifts in the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N dimensions were averaged using the scaling factors for the respective residue type as published in ref *27*. These values have units of parts per million per molar and are in the following termed urea coefficients. The individual $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N urea coefficients can be found in the Supporting Information (Figure S1).

The nonuniform change in secondary chemical shift with an increasing urea concentration was used as a reference-free approach to identify those parts of the sequence that have residual transient structure. Figure 2 shows the urea coefficients plotted as a function of the sequence. The urea coefficients

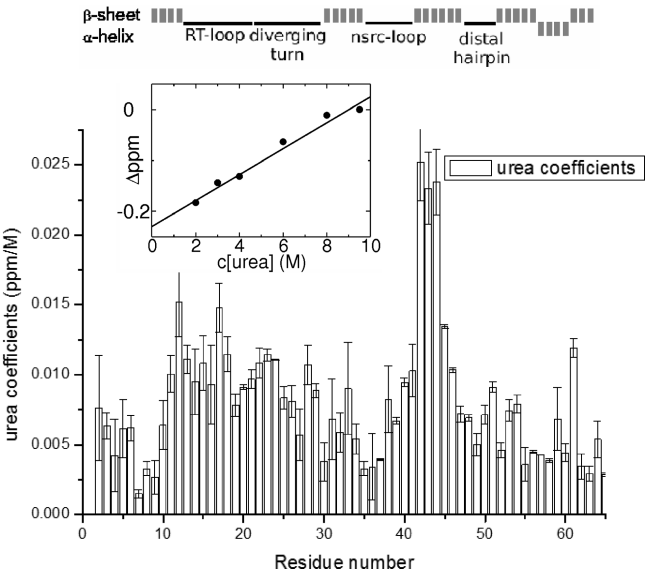

FIGURE 2: Sequence dependence of the combined urea coefficient obtained from the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts at pH 2.0 (see Materials and Methods). The inset shows the urea dependence of the $^{13}C^{\alpha}$ chemical shift of Q52 at pH 2.0, also showing a linear fit as a crude approximation.

displayed one overall maximum between residues 40 and 46 and two smaller maxima between residues 11−18 and 22−29. In the native structure, these regions correspond to the region connecting the n-src loop with the third β-strand and the region between the first and second β-strands, also termed the RT loop, and the diverging turn. Interestingly, the majority of residues experiencing large changes in chemical shift did not correspond to the elements of native secondary structure. In addition, residues identified by the ϕ-value analysis conducted previously (*8*) did not exhibit significant transient secondary structure propensity in the unfolded state.

Secondary Shifts of the Fully Unfolded Protein in the Absence of Urea. The linear regression used to obtain the urea coefficients also yielded a *y*-intercept, which could be used to locate the peak position in the absence of urea. This strongly supported the assignment of the very weak peaks arising from the unfolded state under conditions where the native state dominated the equilibrium. The extrapolated chemical shifts in the absence of urea can be found in the Supporting Information (Table S1). These chemical shifts were analyzed by classical secondary chemical shift analysis. The chemical shifts (δ) in the absence of urea were referenced to random coil shift δ_{rc}. For the analysis, the residue-specific intrinsic random coil chemical shifts (*23*) in a nearly saturated urea solution containing 9.5 M urea (pH 2.0) were used as an "internal" reference set for random coil chemical shifts. Effects of urea on the spectral offset were minimized by referencing the spectra to DSS added to each sample. We found that the properties of the complex peptide chain in the unfolded state are best modeled using this reference set. Alternative reference random coil chemical shifts (*19*−*22*) added a considerable amount of noise to the data. This effect has been described previously (*23*). We could confirm that "external" reference sets model the random coil state with lower accuracy.

The profile of the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N secondary shifts as a function of sequence (see Figure 3) indicates structural preferences in the unfolded ensemble. As expected, none of the secondary shifts was significantly large enough to be proof for a

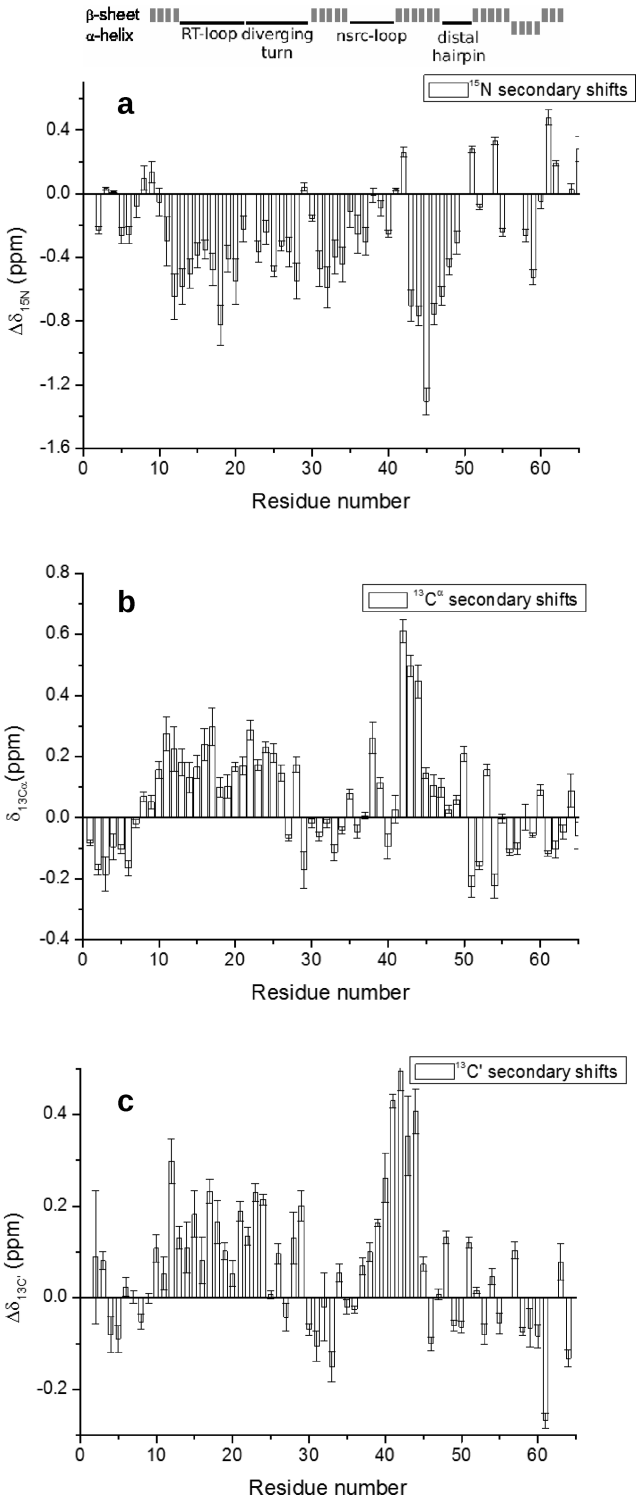

FIGURE 3: Sequence dependence of the secondary chemical shifts of (a) ^{15}N, (b) $^{13}C^{\alpha}$, and (c) $^{13}C'$ of c-src SH3 at pH 2.0 obtained by using protein and site-specific random coil chemical shift references (*23*).

well-defined type of secondary structure. However, the distribution of increased secondary chemical shifts over the sequence was still indicating formation of transient local secondary structure in a considerable number of residues. Figure 3 shows the ^{15}N, $^{13}C^{\alpha}$, and $^{13}C'$ secondary chemical shifts as a function of the sequence. Comparatively large secondary shifts were located between the first and second β-sheet, as well as between residues 40 and 47, i.e., preceding and in the beginning of the third β-sheet. Moreover, the secondary shifts in these two regions were positive for the $^{13}C^{\alpha}$ and $^{13}C'$ chemical shifts and negative for the ^{15}N chemical shifts. This clearly suggested an equilibrium between helical and extended structures. Comparison with the more traditionally used chemical shift referencing values (*19*, *20*, *28*) confirmed the preference of helical structure in the unfolded ensemble of c-src SH3. The analysis of secondary shifts of the unfolded ensemble in the absence and presence of moderate concentrations of denaturant showed a clear trend toward a preference for the helical φ/ψ conformational space when approaching native-like conditions. Most importantly, as one can see in Figure 3, the variations of the secondary shifts of the ^{15}N and the $^{1}H^{N}$ nuclei (the latter not shown) appeared to be shifted approximately three to five residues relative to the position of the $^{13}C^{\alpha}$ and $^{13}C'$ secondary shifts. This trend was previously observed for the acid-unfolded state of ACBP (*23*) and was interpreted as a clear indication of transient helix formation.

Residual Helicity in the Unfolded State of c-src SH3. The percentage of helix formation was estimated using several methods. First, the averaged $^{13}C^{\alpha}$ chemical shifts for helices derived from the BMRB were chosen to act as references for fully formed helices. The internal $^{13}C^{\alpha}$ chemical shifts at 9.5 M urea were chosen as a reference for random coil values. The chemical shifts in the absence of urea were set in reference to these two points, yielding the estimated percentage of helicity (see also Materials and Methods). Negative values should therefore be interpreted as nonhelical but possibly ordered transient structure. The helicity profile is plotted in Figure 4a. This method yielded a maximum helicity of ~45% for residues 43−45 and a maximum helicity of ~9% for residues located in the RT loop and the diverging turn.

The secondary structure propensity score (Ssp) is an unbiased method for secondary shift analysis (*29*). To obtain further verification for the observed trend toward helix formation in the unfolded ensemble, we analyzed chemical shifts of the unfolded state of c-src SH3 at pH 2.0 using this Ssp method (see Figure 4b). This again showed that the backbone of the unfolded state sampled the helical φ/ψ conformational space for residues 40−45. The helicity is estimated by the Ssp method to a maximum value of approximately 20%. The helicity previously found in the RT loop and the diverging turn was not detected using the Ssp method. Instead, the Ssp method predicts a random coil structure for this region. In contrast to the first method applied, the Ssp method applies the reference value of −1 to 100% β-sheet structure. This might explain the slightly different results with respect to the helicity obtained from our first approach, which is using the intrinsic random coil chemical shifts as a reference for 0% helicity instead.

Both helicity profiles were compared with the helical content predicted by AGADIR (*30*) at pH 2.0 and an ionic strength of 0.25 M. Despite the difference of one order of magnitude, there is significant agreement about the location of the helical regions between the experimental and theoretical results (see Figure 4c). Alternative helix prediction algorithms such as GOR4, SIMPA96, and PREDATOR appear to estimate the helicity to up to 30 times larger than the respective AGADIR prediction in a study of hNck2 SH3 (*17*). However, these algorithms were omitted in our study because of the substantial differences from our experimental acidic unfolding conditions. Interestingly, when comparing the AGADIR helicity prediction of c-src SH3 to those of various other SH3 domains, we noticed that the locations of increased helical propensity under these conditions are not necessarily conserved throughout different members of the SH3

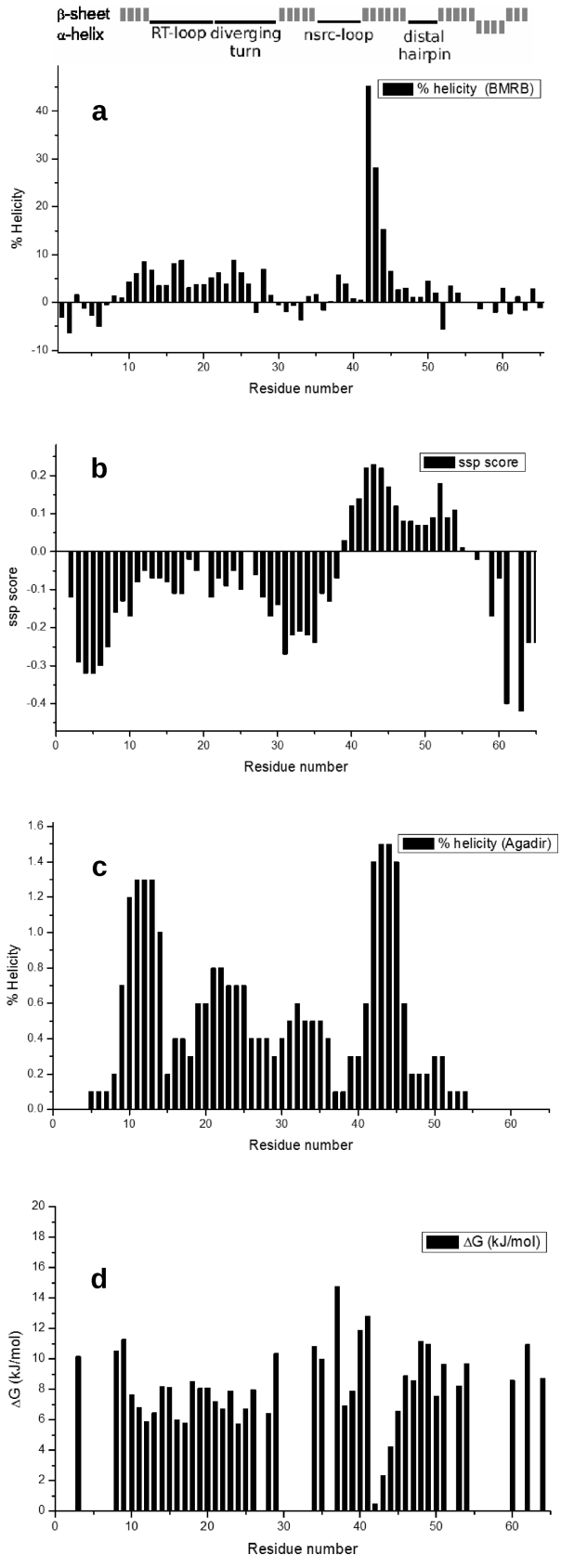

FIGURE 4: (a) Helicity as obtained from the $^{13}C^{\alpha}$ chemical shifts at pH 2.0 compared with the chemical shifts for fully formed α helices derived from the BMRB. (b) Ssp analysis of the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts at pH 2.0 (*29*). (c) Helicity at pH 2.0 and an ionic strength of 250 mM as predicted by AGADIR (*30*). (d) Sequence dependence of ΔG for the $i + 4$ peptide backbone hydrogen bond formation for c-src SH3 at pH 2.0 in the absence of urea based on the $^{13}C^{\alpha}$ secondary chemical shift.

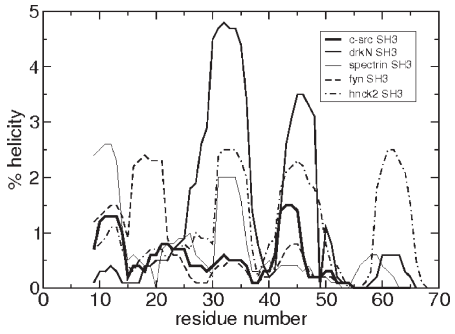

FIGURE 5: Residue-specific helicity prediction by AGADIR (*30*) for c-src SH3 (thick solid line), drkN SH3 (medium solid line), spectrin SH3 (thin solid line), fyn SH3 (dashed line), and hNck2 SH3 (dashed−dotted line). The profile starts with the residues corresponding to T9 of c-src SH3.

fold (see Figure 5). Even though this has no apparent influence on the native state topology, this might have an impact on the stability and dynamics of numerous SH3 domains.

As reported previously (*23*), the residue-specific free energy (ΔG) can be used to investigate whether there are any indications that the backbone might form $i + 3/i + 4$ hydrogen bonds. In Figure 4d, the free energy ΔG is plotted for the unfolded state of c-src SH3. The free energy ΔG shows distinct local minima in the regions of enlarged secondary chemical shifts, indicating a certain degree of $(i + 3)$ or $(i + 4)$ hydrogen bond formation. Thus, we could conclude that the unfolded ensemble contains non-native-like conformations of c-src SH3 that show a tendency to adopt hydrogen bonds between residues i and $(i + 3)/(i + 4)$ in the two regions of V11−K28 and G40−H46.

DISCUSSION

The unfolded ensemble of all-β protein c-src SH3 was studied using secondary chemical shift analysis. Similar studies conducted with ACBP (*3*, *23*) or protein L (*5*) identified transient structures in the denatured ensemble that clearly resembled the secondary structure composition of the TSE. Investigating the degree to which residual structure in the unfolded ensemble would also resemble the native-like TSE as previously described by an extensive ϕ-value analysis (*8*) was worthwhile. We found that the residues being engaged in residual structure in the unfolded ensemble were situated in the regions corresponding to parts of the first β-strand, the RT loop, the diverging turn, and the n-src loop in the native state. Surprisingly, the majority of these residues were situated outside the regions of ordered secondary structure in the native state or TSE. Furthermore, we discovered that their backbone connectivity more closely resembled an α-helix than a β-strand. The comparison with various alternative chemical shift reference sets confirmed that the regions we found appeared to adopt the helical range of φ/ψ conformational space. The comparatively large deviations in the third β-sheet correspond to a local sequence that contains a considerable amount of aromatic residues (^{41}DWWLAH^{46}). We have good reasons to believe that our observations do not arise from ring current effects but indeed do indicate residual helicity in the unfolded ensemble. Most importantly, ring currents are only known to have a marked effect on amide proton and ^{15}N chemical shifts, while especially $^{13}C^{\alpha}$ chemical shifts largely remain unaffected (*31*, *32*). Comparison to the AGADIR prediction at low pH also confirmed that the c-src SH3 can theoretically sample the helical φ/ψ conformational space in the regions identified.

In addition, the tendency to form a helical folding intermediate, not exclusively, but especially under acidic conditions is well-established (*16*). Furthermore, a helical folding intermediate was recently isolated from hNck2 SH3 (*17*).

Transient helix formation has also been reported previously for other all-β proteins such as β-lactoglobulin (*33*) and ubiquitin (*34*). It becomes increasingly apparent that transient α helical conformations could be a general feature on more protein energy landscapes than originally anticipated. Transient helices are relatively easy to form and do not require extensive tertiary contacts. According to our urea coefficient analysis, the regions of residual local order seemed to persist at urea concentrations reaching ≥4 M at pH 2.0. We therefore propose from our results that these transient helices exhibit a clear tendency to be readily sampled in the unfolded ensemble even under very stringent denaturing conditions.

It is of interest to note that almost none of the key residues involved in the initial condensation of the structure could be identified by classical ϕ-value analysis (*8*). Instead, we observed that two separate groups of residues are responsible for the formation of transient structure in the unfolded ensemble and the formation of the TSE of c-src SH3. In the unfolded ensemble, the formation of local structures, such as hydrogen bond and loop/helix initiation, dominates over extensive tertiary contacts observed during the transition to the native state. We can speculate that residues initially engaged in the formation of an alternative connectivity can less easily interfere with the later formation of a hydrophobic core. This might lead to the proposal of alternative strategies for minimizing frustration along the folding pathway.

Non-Native Secondary Structure Contrasts with the Conserved Contacts in the TSE and the Unfolded Ensembles Generated by MD Simulations. Previous molecular dynamics simulations of the unfolding reaction claim that the most persistent native structure in the unfolding kinetics is found in the region surrounding the β-sheet connected by the n-src loop, the diverging turn, and the distal hairpin. Further residual native-like structure appears to be located around the RT loop, especially involving residue D23 and its local hydrogen bond network (*35*). Refolding molecular dynamics simulations clearly identify the RT loop and the n-src loop as the first to emerge on the protein folding pathway. This initial nucleation event is followed by formation of the distal hairpin indicated by the formation of contacts between strands 3 and 4 (*36*). In the Rosetta MD refolding simulation (*8*), the RT loop also makes numerous contacts despite having a low ϕ-value.

Indeed, our chemical shift analysis confirms residual structure in the RT loop region. In addition, our data showed a clear preference for non-native-like helical secondary structure. However, frequently postulated native-like structure in the β-sheet formed by the n-src loop, the diverging turn, and the distal hairpin could not be confirmed. Hence, we suggest that the native-like character is a feature previously overestimated by unfolding MD simulations. In contrast to previous assumptions, we propose a non-native hydrophobic cluster adopting a backbone connectivity dominated by conformations sampling the helical φ/ψ conformational space in the n-src loop and the third β-strand.

Residual Structure in the Unfolded State Ensemble Is Less Conserved among Different Members of the SH3 Family. SH3 domains have been subjected to intense studies with regard to their protein folding pathways and non-native states. In particular, the unfolded state of *Drosophila* drkN SH3 has been characterized in detail because the unfolded state is readily accessible under nearly physiological conditions at pH 6.0. According to AGADIR analysis of the secondary structure profile at pH 6.0 and 2.0, the intrinsic helicity is only marginally affected by pH. The ensemble of unfolded drkN SH3 structures at pH 6.0 shows some similarities but also significant differences with respect to that of c-src SH3. Residual β-strand propensity could be found for residues 5−9 in drkN SH3 (the first β-strand) but not in c-src SH3. c-src SH3 clearly shows a preference for the helical φ/ψ conformational space in the corresponding region. Regions of residual helicity are also described for the drkN SH3 domain but appear to be localized between the first two β-strands of the protein. Similar to c-src SH3, the first half of the RT loop clearly shows a helical propensity (residues 16−20). However, the NMR cross signals of drkN SH3 are broadened beyond detection in the second half of the RT loop (residues 23−29), hampering the direct comparison (*12*). Residue W36 of drkN SH3 corresponds to W41 in c-src SH3. W36 exhibits a considerable non-native burial in the U_{exch} state of drkN SH3 (*37*). Using the ensemble program, numerous additional native-like tertiary contacts are described for drkN SH3 in the same region (*13*). However, these residues show no sign of helical backbone conformations in drkN SH3 which is in clear contrast with the significant helicity seen in the corresponding region in c-src SH3.

Another SH3 domain whose secondary chemical shifts suggest regions of transient helicity in the unfolded ensemble is the AE-K mutant of spectrin SH3. The transient helicity in the unfolded ensemble is situated around residues 3−14, corresponding to the first β-strand and the first half of the RT loop. To some extent, residual helicity can also be observed for residues 30−34 in the n-src loop (*38*). No long-range contacts of the artificially induced helix are described. In contrast to the AE-K mutant, the wild-type protein exhibits no transient helicity but a considerable amount of long-range interactions in the unfolded ensemble. Here, residues 13−34 appear to be in contact with residues 41, 42, and 53, indicating a native-like tertiary connectivity (*39*). Both the mutant and the wild type have similar refolding rates which suggests that the presence or absence of this transient helicity has no apparent impact on the folding pathway (*38*).

The third well-characterized SH3 domain, fyn SH3, also displays no relevant residual secondary structure in the unfolded state (*40*). However, a molten globule-like folding intermediate is proposed, which is comprised of a flexible yet native-like hydrophobic core. In the unfolded state itself, only one non-native-like interaction has been identified. These contacts from residue 53 to position 40 or 47 in the U state are believed to decelerate the folding process (*41*).

When comparing the AGADIR helicity profile of c-src SH3 with the profiles of drkN SH3, hNck2 SH3, fyn SH3, and spectrin SH3 under our experimental conditions (pH 2 and ionic strength of 0.25 M), we could indeed also observe a certain heterogeneity (see Figure 5; the clustalW sequence alignment can be found in Figure S2 of the Supporting Information). This heterogeneity in the unfolded ensembles of various SH3 domains stands in clear contrast to their otherwise highly conserved TSE and native state topology. The AGADIR predictions were found to be in good agreement with the data reported by the various groups. Generally, a certain degree of helicity appears to be more or less conserved in the RT loop. However, the order of magnitude apparently does not interfere with the formation of TSEs and native states as extensive folding studies on various SH3 domains

have demonstrated (*8*−*10*). Other non-native contacts appear to be even less conserved among SH3 domains. Frequently, aromatic residues in the n-src loop appear to be involved in the formation of non-native residual structure, but the degree of reported secondary structure content varies considerably. Furthermore, no obvious correlation between intrinsic helicity and sequence conservation could be detected. It remains of great interest to investigate whether the heterogeneity in the unfolded state might be somehow related to the heterogeneity of the overall dynamics and stability despite the highly conserved TSEs and native structures of all the different members of the SH3 family.

ACKNOWLEDGMENT

We thank Prof. David Shalloway for providing the pGEX-SH3 plasmid containing the coding sequence for c-src SH3.

SUPPORTING INFORMATION AVAILABLE

Individual urea coefficients for the $^{13}C^{\alpha}$, $^{13}C'$, $^{1}H^{N}$, and ^{15}N chemical shifts, a table of the extrapolated chemical shifts in the absence of urea, and the clustalW sequence alignment of the five SH3 domains discussed here. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Eliezer, D. (2007) Characterizing residual structure in disordered protein states using nuclear magnetic resonance. *Methods Mol. Biol. 350*, 49–67.
2. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry 40*, 3561–3571.
3. Teilum, K., Kragelund, B. B., and Poulsen, F. M. (2002) Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. *J. Mol. Biol. 324*, 349–357.
4. Le Duff, C. S., Whittaker, S. B., Radford, S. E., and Moore, G. R. (2006) Characterisation of the conformational properties of urea-unfolded Im7: Implications for the early stages of protein folding. *J. Mol. Biol. 364*, 824–835.
5. Yi, Q., Scalley-Kim, M. L., Alm, E. J., and Baker, D. (2000) NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol. 299*, 1341–1351.
6. Mayer, B. J. (2001) SH3 domains: Complexity in moderation. *J. Cell Sci. 114*, 1253–1263.
7. Noble, M. E., Musacchio, A., Saraste, M., Courtneidge, S. A., and Wierenga, R. K. (1993) Crystal structure of the SH3 domain in human Fyn: Comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J. 12*, 2617–2624.
8. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol. 6*, 1016–1024.
9. Martinez, J. C., and Serrano, L. (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol. 6*, 1010–1016.
10. Northey, J. G., Di Nardo, A. A., and Davidson, A. R. (2002) Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol. 9*, 126–130.
11. Wales, T. E., and Engen, J. R. (2006) Partial unfolding of diverse SH3 domains on a wide timescale. *J. Mol. Biol. 357*, 1592–1604.
12. Zhang, O., and Forman-Kay, J. D. (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry 36*, 3959–3970.
13. Marsh, J. A., Neale, C., Jack, F. E., Choy, W. Y., Lee, A. Y., Crowhurst, K. A., and Forman-Kay, J. D. (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J. Mol. Biol. 367*, 1494–1510.
14. Marsh, J. A., and Forman-Kay, J. D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol. 391*, 359–374.
15. Chikenji, G., Fujitsuka, Y., and Takada, S. (2004) Protein folding mechanisms and energy landscape of src SH3 domain studied by a structure prediction toolbox. *Chem. Phys. 307*, 157–162.
16. Li, J., Shinjo, M., Matsumura, Y., Morita, M., Baker, D., Ikeguchi, M., and Kihara, H. (2007) An α-helical burst in the src SH3 folding pathway. *Biochemistry 46*, 5072–5082.
17. Liu, J., and Song, J. (2008) NMR evidence for forming highly populated helical conformations in the partially folded hNck2 SH3 domain. *Biophys. J. 95*, 4803–4812.
18. Li, J., Matsumura, Y., Shinjo, M., Kojima, M., and Kihara, H. (2007) A stable α-helix-rich intermediate is formed by a single mutation of the β-sheet protein, src SH3, at pH 3. *J. Mol. Biol. 372*, 747–755.
19. Schwarzinger, S., Kroon, G. J., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift data in NMRView. *J. Biomol. NMR 18*, 43–48.
20. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc. 123*, 2970–2978.
21. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) H-1, C-13 and N-15 Random Coil NMR Chemical Shifts of the Common Amino Acids. 1. Investigations of Nearest-Neighbor Effects. *J. Biomol. NMR 5*, 332.
22. Braun, D., Wider, G., and Wüthrich, K. (1994) Sequence-corrected ^{15}N "random coil" chemical shifts. *J. Am. Chem. Soc. 116*, 8466–8469.
23. Modig, K., Jurgensen, V. W., Lindorff-Larsen, K., Fieber, W., Bohr, H. G., and Poulsen, F. M. (2007) Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. *FEBS Lett. 581*, 4965–4971.
24. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q., and Baker, D. (1997) Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol. 4*, 805–809.
25. Grantcharova, V. P., and Baker, D. (1997) Folding dynamics of the src SH3 domain. *Biochemistry 36*, 15685–15692.
26. Berjanskii, M. V., and Wishart, D. S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc. 127*, 14970–14971.
27. Wishart, D. S., Arndt, D., Berjanskii, M., Tang, P., Zhou, J., and Lin, G. (2008) CS23D: A web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res. 36*, W496–W502.
28. Wang, Y., and Jardetzky, O. (2002) Investigation of the neighboring residue effects on protein chemical shifts. *J. Am. Chem. Soc. 124*, 14075–14084.
29. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci. 15*, 2795–2804.
30. Munoz, V., and Serrano, L. (1997) Development of the multiple sequence approximation within the AGADIR model of α-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers 41*, 495–509.
31. Blanchard, L., Hunter, C. N., and Williamson, M. P. (1997) The effect of ring currents on carbon chemical shifts in cytochromes. *J. Biomol. NMR 9*, 389–395.
32. Iwadate, M., Asakura, T., and Williamson, M. P. (1999) C-α and C-β carbon-13 chemical shifts in proteins from an empirical database. *J. Biomol. NMR 13*, 199–211.
33. Kuwajima, K., Yamaya, H., and Sugai, S. (1996) The burst-phase intermediate in the refolding of β-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy. *J. Mol. Biol. 264*, 806–822.
34. Larios, E., Li, J. S., Schulten, K., Kihara, H., and Gruebele, M. (2004) Multiple probes reveal a native-like intermediate during low-temperature refolding of ubiquitin. *J. Mol. Biol. 340*, 115–125.
35. Tsai, J., Levitt, M., and Baker, D. (1999) Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol. 291*, 215–225.
36. Shea, J. E., Onuchic, J. N., and Brooks, C. L., III (2002) Probing the folding free energy landscape of the Src-SH3 protein domain. *Proc. Natl. Acad. Sci. U.S.A. 99*, 16064–16068.
37. Crowhurst, K. A., Tollinger, M., and Forman-Kay, J. D. (2002) Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. *J. Mol. Biol. 322*, 163–178.
38. Blanco, F. J., Serrano, L., and Forman-Kay, J. D. (1998) High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. *J. Mol. Biol. 284*, 1153–1164.

39. Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J., and Serrano, L. (2000) Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol. 297*, 1217–1229.
40. Mittermaier, A., Korzhnev, D. M., and Kay, L. E. (2005) Side-chain interactions in the folding pathway of a Fyn SH3 domain mutant studied by relaxation dispersion NMR spectroscopy. *Biochemistry 44*, 15430–15436.
41. Zarrine-Afsar, A., Wallin, S., Neculai, A. M., Neudecker, P., Howell, P. L., Davidson, A. R., and Chan, H. S. (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc. Natl. Acad. Sci. U.S.A. 105*, 9999–10004.
42. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr. 24*, 946–950.