NMR Evidence for Forming Highly Populated Helical Conformations in the Partially Folded hNck2 SH3 Domain

Jingxian Liu* and Jianxing Song*†
*Department of Biochemistry, Yong Loo Lin School of Medicine, and †Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore 119260, Singapore

ABSTRACT Recent studies of several proteins implied that the folding of β-proteins may follow a nonhierarchical mechanism in which two major transitions are essential, i.e., the collapse of a random coil to form a nonnative helical intermediate, followed by a transformation into the native β-structure. We report that the first hNck2 SH3 domain, assuming an all-β barrel in the native form, can be reversibly transformed into a stable and nonnative helical state by acid-unfolding. We also conducted extensive NMR and mutagenesis studies that led to two striking findings: 1), NMR analysis reveals that in the helical state formed at pH 2.0, the first and last β-strands in the native form become unstructured, whereas the rest is surprisingly converted into two highly populated helices with a significantly limited backbone motion; and 2), a conserved four-residue sequence is identified on the second β-strand, a mutation of which suddenly renders the SH3 domain into a helical state even at pH 6.5, with NMR conformational and dynamic properties highly similar to those of the wild-type at pH 2.0. This observation implies that the region might contribute key interactions to disrupt the helical state, and to facilitate a further transformation into the native SH3 fold in the second transition.

INTRODUCTION

Many proteins fold into unique three-dimensional structures composed of spatially organized polypeptide fragments that assume different secondary structures, including the α-helix, β-sheet, reverse turns, and loops. Some proteins were found to fold into their native states via intermediates, and in particular, the partially folded states termed ''molten globule'' were studied extensively (1–10). Many models for protein-folding emphasize the hierarchical formation of the native-like substructures during the folding process. On the other hand, recent studies of several proteins, and particularly β-lactoglobulin, implied that the folding of β-proteins may follow a nonhierarchical mechanism in which two major transitions are essential to reach the final native β-structure. More specifically, the first transition is involved in the collapse of the random-coil-like polypeptide chain into a nonnative helical intermediate mainly specified by local interactions, whereas the second transition is associated with transformation into the native β-structure, with the helical conformation disrupted by long-range interactions (11–17). However, the conformational and dynamic properties of such nonnative helical states remain poorly understood. Moreover, the sequence determinants for transforming the helical intermediate into the native β-structure have not been identified.

To date, more than 4,000 SH3 modular domains have been identified in a variety of organisms. The SH3 domains, containing ~60 residues and no disulfide bridge, play a critical role in transmitting as well as integrating cellular signals, mainly by binding to proline-rich short motifs (18–20). Structurally, all SH3 domains share a common β-barrel fold comprising five β-stands, which are organized into two β-sheets. One consists of the last and first β-strands plus the first half of the second β-strand, whereas the other contains the second half of the second β-strand, together with the third and fourth β-strands. The second β-strand seems to play a critical role in coordinating the two β-sheets together, to form the well-known SH3 fold (Fig. 1 *a*).

The 380-residue adaptor protein Nck2 is composed of three SH3 domains and one C-terminal SH2 domain. In vivo, it functions to coordinate the signaling networks critical for the organization of the actin cytoskeleton, cell movement, or axon guidance, by connecting cellular surface receptors down to the multiple intracellular signaling networks in a ''Tyr(P) → SH2/SH3 → effector'' manner (21,22). The NMR structure of the first Nck2 SH3 domain was previously determined (23), and we determined the structures of the second and third SH3 domains as well as the SH2 domain (20,22). Because the SH3 domain has a small all-β fold and contains no disulfide bridge, we are using it as a model system for investigating the folding mechanism of β-proteins. Unexpectedly, we discovered, without needing to introduce any cosolvent or stabilizer, that the wild-type form of the first Nck2 SH3 domain could be reversibly converted into a stable helical state at equilibrium at pH 2.0, as detected by circular dichroism CD spectroscopy. To acquire further insights, we conducted extensive CD and NMR investigations, and the results not only offered an NMR conformational and dynamic view of this nonnative helical state, but also led to the further identification of a four-residue sequence that appears to play a key role in transforming the helical state into the native SH3 fold. When we started preparing the present manuscript, a study was published demonstrating that a nonnative helical intermediate was also populated in the kinetic refolding of the Src SH3 domain (24). Another article from the same group showed

Submitted November 12, 2007, and accepted for publication June 5, 2008.

Address reprint requests to Jianxing Song, Dept. of Biological Sciences, Faculty of Science, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore. Tel.: 65-6874-1013; Fax: 65-6779-2486; E-mail: bchsj@nus.edu.sg.

Editor: Heinrich Roder.

0006-3495/08/11/4803/10 $2.00

doi: 10.1529/biophysj.107.125641

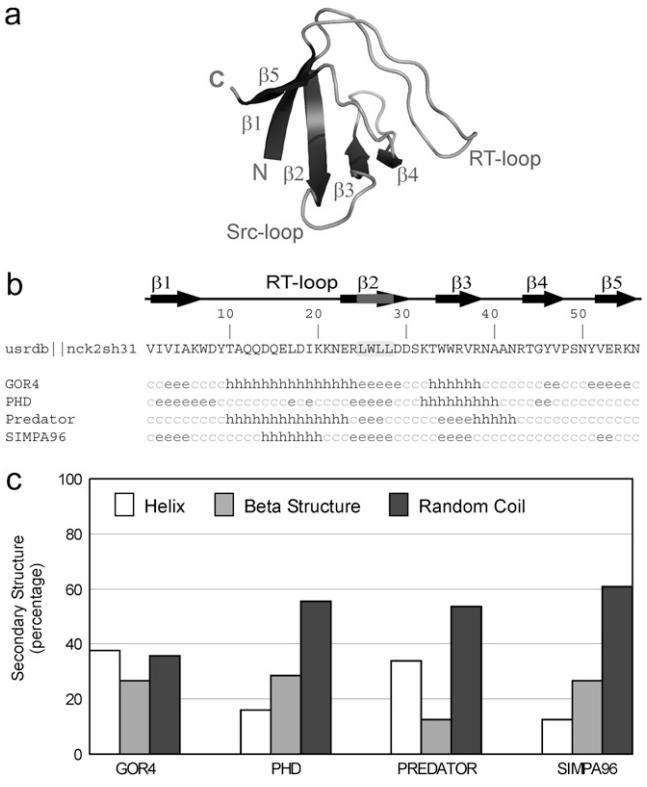

FIGURE 1 Bioinformatics analysis of secondary structures. (*a*) Three-dimensional structure of first hNck2 SH3 domain determined by NMR spectroscopy (23). (*b*) Secondary-structure prediction of first hNck2 SH3 domain by computational programs GOR4, PHD, Predator, and SIMPA96. *e*, β-strand; *h*, helix; *c*, random coil. The four residues ''LWLL'' identified here are highlighted in gray. (*c*) Bar plots for secondary structure contents, as predicted by GOR4, PHD, Predator, and SIMPA96.

that the Ala^{45}-Gly mutant of the Src SH3 domain could also form a stable helical intermediate at equilibrium, as indicated by CD, fluorescence, and x-ray scattering (25).

MATERIALS AND METHODS

Secondary structure prediction

Six computational programs were used to predict the secondary structure of the first hNck2 SH3 domain, i.e., GOR4 (26), PHD (27), Predator (28), SIMPA96 (29), Helix2 (30,31), and AGADIR (32,33).

Cloning, expression, and purification

To achieve high-level protein expression, the DNA fragment encoding the first hNck2 SH3 domain with *Escherichia coli*-preferred codons was obtained by a polymerase chain reaction-based de novo gene synthesis approach, as previously described (8,20), and subsequently cloned into pET32a vector (Novagen, Singapore). Similarly, the DNA fragment for the SH3 domain carrying four-residue Ala-mutations over residues Lue^{25}-Trp^{26}-Leu^{27}-Leu^{28} (designated 4AlaMut) was synthesized by the same approach, except that the two middle primers were replaced by those with the mutations.

The recombinant SH3 domains were overexpressed in *E. coli* strain BL21 cells. Briefly, the cells were cultured at 37°C to reach an optical density at 600 nm of 0.4, and then isopropyl β-D-1-thiogalactopyranoside was added to a final concentration of 0.4 mM to induce the expression of the recombinant protein overnight at 20°C. However, probably because of the partial folding, the expression level of the recombinant 4AlaMut was much lower than that of the wild-type. The recombinant wild-type and 4AlaMut were first purified by a Ni^{2+}-agrose (Qiagen, Singapore) affinity column under native conditions, as recommended by Qiagen, followed by direct high-performance liquid chromatography purifications on an RP-18 column (Vydac, Illinois). The HPLC-purified fractions were lyophilized for further CD and NMR experiments.

For NMR isotope labeling, recombinant proteins were prepared by growing the cells in M9 medium with additions of $(^{15}NH_4)_2SO_4$ for ^{15}N labeling, and $(^{15}NH_4)_2SO_4$ and [^{13}C]-glucose for double-labeling, respectively (20). The identities of proteins were verified by matrix assisted laser desorption/ionization time-of-flight mass spectrometry.

CD and NMR experiments

All CD measurements were conducted on a Jasco J-810 spectropolarimeter (JASCO, Easton, MD) equipped with a thermal controller, using a 1-mm path-length cuvette with 0.1-nm spectral resolution. Far-ultraviolet (UV) spectra were recorded in the range of 190–260 nm at different pH values, and data from three independent scans were added and averaged. The sample concentrations for far-UV CD measurements ranged from 5–100 μM in 5 mM phosphate buffer. The secondary structure was analyzed by the CDPro software package (http://lamar.colostate.edu/~sreeram/CDPro).

The NMR samples for both the wild-type and 4AlaMut were prepared by dissolving the protein powders into 5 mM phosphate buffer, with a final pH of 6.5. The wild-type samples at pH 2.0 were obtained by adding diluted hydrochloric acid. All these samples were concentrated to desired concentrations by Centricon (Millipore, Singapore), with a 3-kDa threshold. All NMR experiments were performed on an 800-MHz Bruker Avance spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) equipped with pulse-field gradient units at 278 K, as described previously (20,34). For the wild-type at pH 6.5, only a pair of triple-resonance experiments HNCACB and CBCA(CO)NH was acquired for the sequential assignment. For the wild-type at pH 2.0, extensive NMR data were collected, including two-dimensional ^{1}H-TOCSY and NOESY (mixing times of 250 ms and 350 ms), and three-dimensional heteronuclear ^{1}H-^{15}N HSQC (heteronuclear single quantum coherence)-TOCSY (total correlation spectroscopy), HSQC-NOESY (nuclear Overhauser enhancement Spectroscopy), HNCACB (interresidual correlations), CBCA(CO)NH (interresidual correlations), HCCH (interresidual correlations)-TOCSY, and HCCH-NOESY. All above-mentioned data were collected for 4AlaMut, except in the triple-resonance experiments. To assess whether significant aggregation existed, HSQC spectra were also collected, in a series of protein concentrations ranging from 50–1,000 μM.

The NMR ^{15}N-relaxation data were collected using the same 800 MHz Bruker Avance spectrometer for the wild-type at both pH 6.5 and 2.0, and for 4AlaMut at pH 6.5 by the method previously described (20,35). The ^{15}N $T1$ values for the wild-type at pH 6.5 were obtained by fitting HSQC spectra recorded with relaxation delays of 10, 400, 100, 300, 200, 350, and 250 ms, whereas ^{15}N $T2$ values were recorded with relaxation delays of 10, 30, 45, 60, 75, 90, and 150 ms. For the wild-type at pH 2.0, relaxation delays for ^{15}N $T1$ were 10, 700, 100, 600, 200, 500, 300, and 400 ms; and for ^{15}N $T2$, 10, 60, 100, 130, 160, 200, 230, and 260 ms. For 4AlaMut, the relaxation delays for ^{15}N $T1$ were 10, 350, 50, 300, 100, 400, and 200 ms, whereas for ^{15}N $T2$, they were 10, 30, 60, 90, 120, 140, and 160 ms. The {^{1}H}-^{15}N steady-state NOEs were obtained by recording spectra with and without ^{1}H presaturation at a duration of 3 s, plus a relaxation delay of 6 s at 800 MHz. Spectra processing and analysis were performed using NMRpipe (36) and NMRview (37) software. The structural models in Fig. 7 were generated by the graphic software Pymol (www.pymol.org). The native NMR structures of the first hNck2 and Src SH3 domains were obtained from the Protein Data Bank, with codes of 2B86 and 1SRL, respectively.

RESULTS

Bioinformatics and CD characterization

Previously, a high-resolution NMR structure was reported for the first human Nck2 SH3 domain (23), which unambiguously

showed that it adopts the classic all-β fold common to all SH3 domains (Fig. 1 *a*). On the other hand, as seen in Fig. 1, *b* and *c*, prediction of its secondary structures by several computational programs, including GOR4 (26), PHD (27), Predator (28), and SHIMPA96 (29), suggested that a large portion of the SH3 domain possessed a high intrinsic propensity to form helical conformations. In particular, the results of GOR4 and Predator indicated that the SH3 domain had an even higher percentage of helix than did the β-sheet (Fig. 1 *c*). However, because these programs predict protein secondary structure with the incorporation of empirical information including the tertiary structure from the Protein Data Bank (24), we used Helix2, which implements the Lifson-Roig helix-coil transition theory and only uses the interaction between adjacent residues (24,30,31). The results also showed that the SH3 domain contained a 21% helix fraction. However, the prediction by AGADIR (32,33) only yielded 2.2% and 1.1% helix fractions at pH 6.5 and 2.0, respectively. In general, it is widely accepted that helix formation is mainly driven by local interactions (30–33), whereas β-sheet formation largely depends on long-range interactions and thus is context-dependent (38,39). As such, the observed inconsistency between the NMR structure and prediction results inspired us to speculate that at least for the first hNck2 SH3 domain, the formation of the all-β SH3 native fold might be extensively driven by specific long-range interactions, which also function to override the intrinsic helix-formation propensity. Therefore, destabilization of the tertiary packing would trigger a conversion of the all-β SH3 fold into a helical state.

As shown in Fig. 2, at pH 6.5, the wild-type SH3 domain had a far-UV CD spectrum with some unique characteristics (40). The large negative signal at 203 nm and the positive signal at 227 nm suggest the presence of some polyproline II structure in the native state of the protein, although there is only one proline residue in the SH3 domain. Although all SH3 domains have the same three-dimensional fold, they have diverse far-UV CD spectra and dynamic properties, as

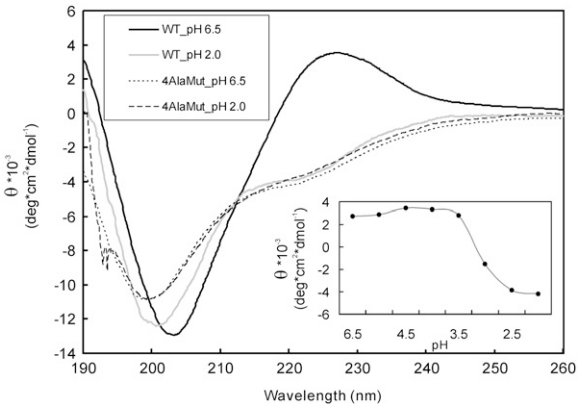

FIGURE 2 Far-UV CD characterization of first hNck2 SH3 domain: far-UV CD spectra for wild-type at pH 6.5 (*black*), wild-type at pH 2.0 (*gray*), 4AlaMut at pH 6.5 (*dotted line*), and 4AlaMut at pH 2.0 (*dashed line*). (*Inset*) The pH-induced conformational changes of the wild-type as monitored by ellipticity at 222 nm.

previously documented (20,41). The origin of these unique properties is not well-understood, and some might result from the existence of the very long (>10 residues) and unique RT-loop in all SH3 domains, which adopts no regular secondary structure, but makes extensive tertiary contacts with other parts of the SH domains (20). By contrast, it appeared that at pH 2.0, the SH3 domain switched into a partially folded state containing some residual helical conformation. Analysis of the CD spectrum by different programs gave variable secondary-structure fractions: 15% helix, 35% turn/strand, and 5% random coil by CONTINLL; 12% helix, 46% turn/strand, and 42% random coil by SELCON3; and 5% helix, 59% turn/strand, and 36% random coil by CDSSTR (42). Furthermore, the fraction of helix was estimated at ~11% by assuming the coexistence of the helix and random coil (24). Detailed CD investigations also revealed that the pH-induced conformational change was reversible, with the transition occurring from pH 4 to 2, as monitored at 222 nm (Fig. 2, *inset*).

We subsequently conducted extensive NMR studies. The results confirmed (consistent with the CD observations) that the first Nck2 SH3 domain switched into a highly populated helical state at pH 2.0. This result raised the question of whether a sequence region could be identified in the first hNck2 SH3 domain that was critical for transforming the helical state into the native SH3 fold. In other words, mutation of this region would be anticipated to trap the SH3 domain permanently in the helical state. To address this question, we aligned a large number of SH3 sequences selected from all subfamilies, and the results led to the identification of a four-residue region, Leu^{25}-Trp^{26}-Leu^{27}-Leu^{28}, on the second β-strand (Fig. 1 *b*, *boxes*). Three of these residues were highly conserved in the majority of sequences (Fig. S1 in Supplementary Material, Data S1). Further prediction of secondary structures of a variety of SH3 sequences by different programs also showed that this short region always formed a β-strand (data not shown). Therefore, we experimentally constructed a mutant with these four residues replaced by Ala residues. Surprisingly, as seen in Fig. 2, the 4AlaMut protein had very similar CD spectra at pH values of 6.5 and 2.0, revealing that it completely lost the ability to undergo the pH-induced conformational changes observed in the wild-type SH3 domain. Importantly, even at pH 6.5, 4AlaMut had a far-UV CD spectrum similar to that of the wild-type at pH 2.0, indicating that 4AlaMut formed a partially folded state bearing a similar fraction of the helical conformation even at a neutral pH.

Far-UV CD spectra were also acquired at different protein concentrations ranging from 5–100 μM for the wild-type and 4AlaMut. Superimposition of the spectra showed that they were concentration-independent, indicating that no significant aggregation occurs under these concentrations.

NMR characterization

Isotope-labeled recombinant wild-type hNck2 SH3 and 4AlaMut were prepared for a detailed characterization by

NMR. No significant changes in crosspeak width and position were detected in their HSQC spectra, acquired at different concentrations (50–1,000 μM), indicating that the protein is monomeric under these experimental conditions. Analysis of the triple-resonance NMR spectra led to successful sequential assignments for the wild-type SH3 domain at both pH 6.5 and 2.0 (Table S2 in Data S1). As shown in Fig. 3 *a*, the wild-type SH3 domain had an HSQC spectrum typical of a well-folded β-protein, with spectral dispersions of ~2.7 ppm for the amide ^{1}H and ~22 ppm for the ^{15}N dimensions. By contrast, the HSQC spectral dispersion was significantly reduced (~0.9 ppm for the amide ^{1}H, and ~18 ppm for ^{15}N) for the wild-type at pH 2.0 (Fig. 3 *b*), and for 4AlaMut at pH 6.5 (Fig. 3 *c*), indicating that their tight tertiary packing was severely disrupted. Thus, they are only partially folded or unfolded under these experimental conditions. For comparison, we downloaded the chemical shift deposit (BMRB accession number 6854) previously used for the structural determination of the first hNck2 SH3 domain (23). Although our hNck2 SH3 construct is four residues shorter at the N-terminus and one residue longer at the C-terminus, the Cα and Cβ chemical shifts of the majority of residues at pH 6.5 obtained here are almost identical to those previously deposited, indicating that at pH 6.5, the first hNck2 SH3 domain we studied here adopts the same solution structure as previously determined (23). Because NMR chemical-shift deviations from those expected for random coils ($\Delta\delta = \delta$obs $-$ δcoil) are very sensitive indicators of protein secondary structure (43–46), we calculated Cα and Hα chemical-shift deviations from the random-coil values previously reported (43,44). As clearly indicated by the Cα deviation (Fig. 4 *a*), the wild-type SH3 domain had a β-like secondary structure at pH 6.5, but surprisingly, upon lowering the pH value to 2.0, it switched into a helical conformation. More specifically, it appeared that this β-to-helix transition occurred over the majority of the RT-loop, i.e., the second and third β-strands. The large Hα deviations in Fig. 4 *b* confirmed the formation of highly populated helical fragments with a similar pattern in both the wild-type at pH 2.0 and 4AlaMut at pH 6.5.

The NMR conformational shift was extensively used to estimate the population of the secondary structure in partially folded peptides and proteins (47,48). Recently, it was shown

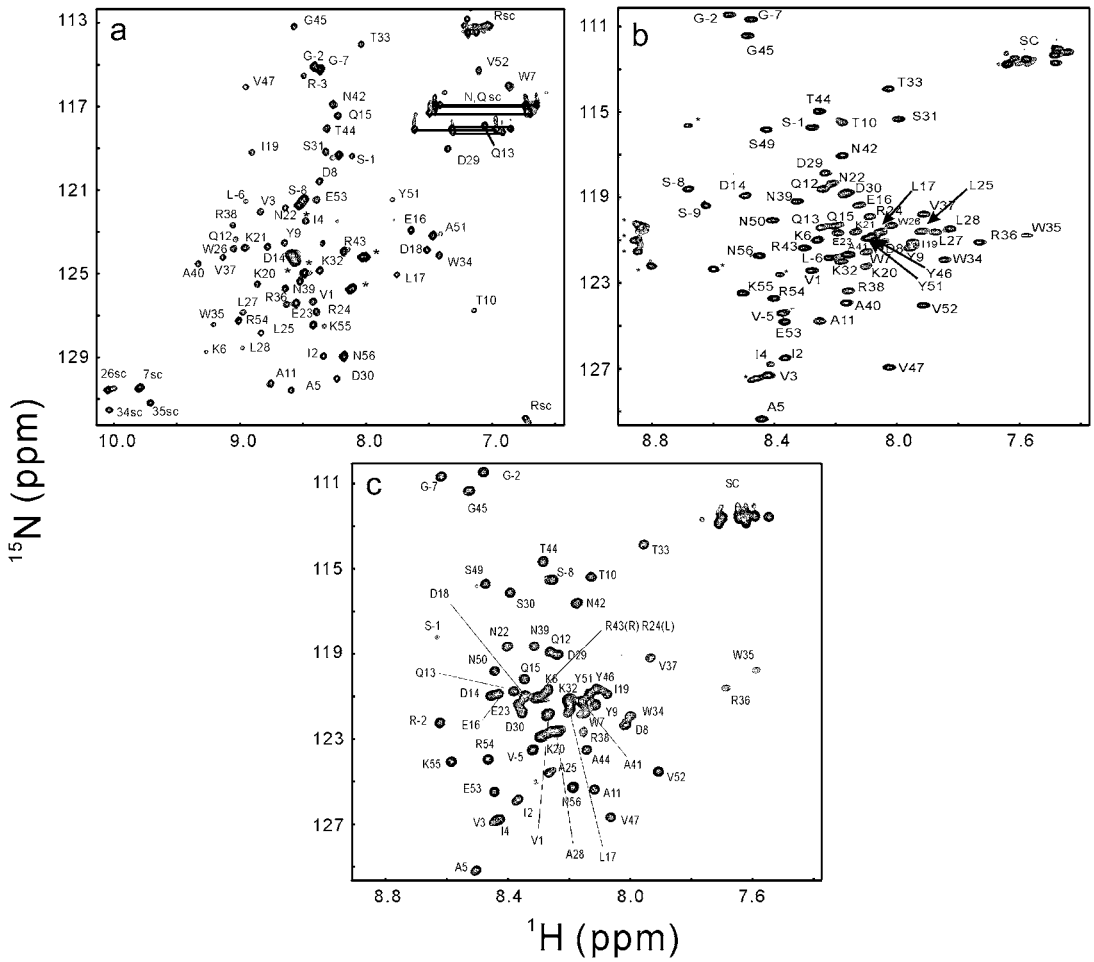

FIGURE 3 The ^{1}H-^{15}N NMR HSQC spectra for different forms of first hNck2 SH3 domain. (*a*) Wild-type at pH 6.5. (*b*) Wild-type at pH 2.0. (*c*) 4AlaMut at pH 6.5. All spectra were acquired on an 800-MHz NMR spectrometer at 278 K. Sequential assignments were labeled for all spectra.

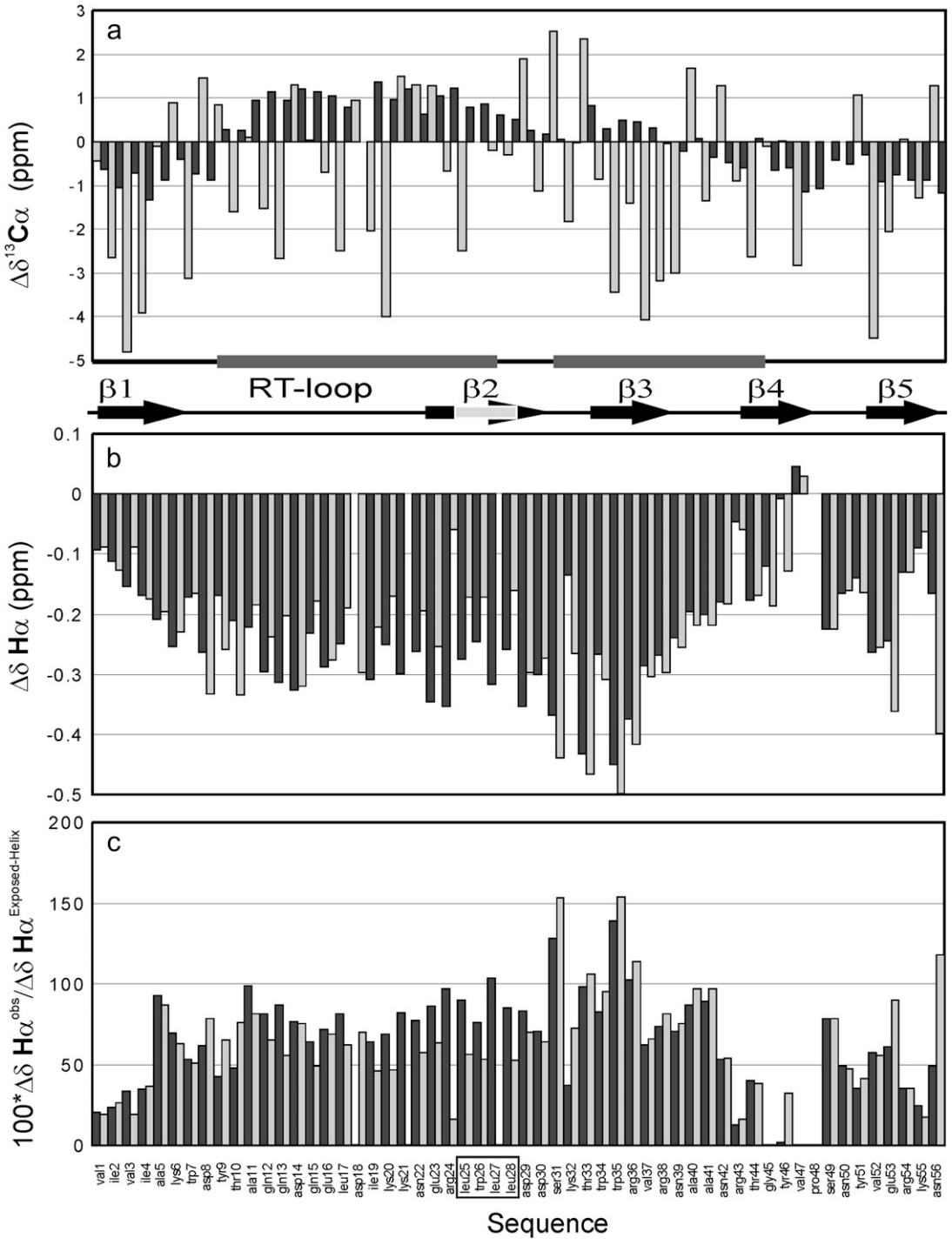

FIGURE 4 Chemical-shift deviations and helix populations. (*a*) Bar plot of Cα chemical-shift deviations from their random-coil values for wild-type at pH 6.5 (*gray*) and pH 2.0 (*black*). (*b*) Bar plot of Hα chemical-shift deviations for wild-type at pH 2.0 (*black*) and 4AlaMut at pH 6.5 (*gray*). (*c*) Ratio (× 100) between Hα deviations of wild-type at pH 2.0 (*black*), 4AlaMut at pH 6.5 (*gray*), and fully folded but exposed helix residues previously reported (50). Black box indicates mutation region. Secondary-structure fragments are also indicated.

that conformational shifts of amino acids were highly dependent on the extent of solvent exposure (46). Statistically, the chemical-shift deviations of residues exposed to solvent are smaller than those of buried residues in the same secondary structure. As judged according to their narrow HSQC dispersions, both the wild-type SH3 domain at pH 2.0 and 4AlaMut at pH 6.5 lacked tight tertiary packing, and as a result, the majority of residues were anticipated to be exposed to the bulk solvent. Therefore, it is reasonable to use average chemical shifts for the exposed helix residues (46) as a ref-

erence for the fully formed but exposed helix. In this regard, by assuming a two-state (random coil and helix) folding model, the helix populations can be approximately represented by the ratio between the chemical-shift deviations observed for the two SH3 forms and those for the fully folded but exposed helix residues previously reported (46). As seen in Fig. 4 *c*, similar helical conformations were populated in the wild-type at pH 2.0 and in 4AlaMut at pH 6.5, except that in 4AlaMut, the helix populations of residues close to the mutation sites were lower than those of corresponding residues in the wild-type at pH 2.0. For the wild-type at pH 2.0, ~70% of the residues showed >50% populated helical conformation, and more than half of the residues had >70% populated helical conformation. Interestingly, residues Ser^{31}, Trp^{35}, and Arg^{36} had conformational shifts even higher than those calculated from the average chemical shifts expected for the exposed but fully folded helix. These results imply that, except for the N-termini and C-termini, helical conformations are highly populated in both the wild-type at pH 2.0 and 4AlaMut at pH 6.5.

Consistent with the chemical shift-based analysis, the NOE patterns defining secondary structures (49) again support the notion that helical conformations with a similar pattern were populated in the wild-type at pH 2.0 as well as in 4AlaMut at pH 6.5. In more detail, as judged according to the characteristic NOE connectivities in Fig. 5 such as $d_{\alpha N}$ (i, i + 2), $d_{\alpha N}$ (i, i + 3), $d_{\alpha N}$ (i, i + 4), and d_{NN} (i, i + 2), it appears that in the two nonnative helical states, the N-termini and C-termini were unstructured, whereas the remaining regions were transformed into two helices linked by a less structured loop over residues Leu^{28}-Asp^{29}-Asp^{30}. However, as judged according to the $d_{\alpha N}$ (i, i + 2) NOEs and the very limited numbers of $d_{\alpha N}$ (i, i + 4) NOEs, the helices observed here might be mainly 3_{10}-helices, which are rarely seen in native proteins but are regarded as a dynamic helix form existing in partially folded proteins.

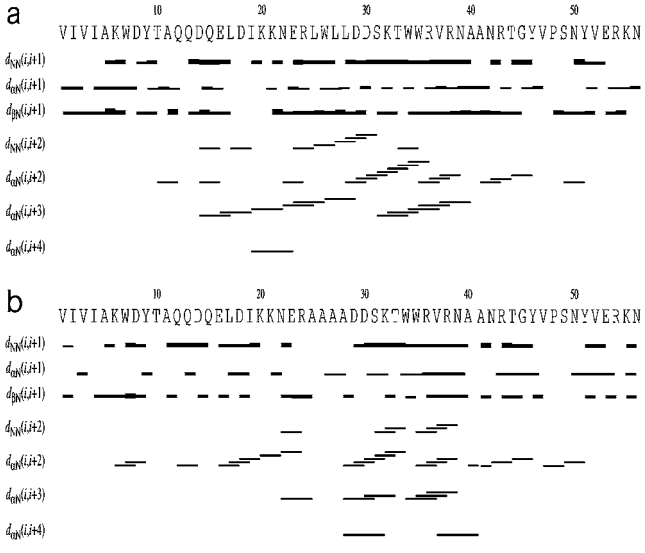

FIGURE 5 Characteristic NOEs defining secondary-structure NOE connectivities identified for (*a*) wild-type at pH 2.0 and (*b*) 4AlaMut at pH 6.5. Plots were generated by CYANA 2.1 (Güntert, riken, Wako, Saitama, Japan).

NMR ^{15}N backbone relaxation

The ^{15}N NMR backbone relaxation data provide valuable information about the dynamics of the local environment of a protein on the picosecond-to-nanosecond timescale (35,50). In particular, the $\{^1H\}$-^{15}N heteronuclear steady-state NOE (hNOE) provides a measure of the backbone flexibility. Interestingly, in a large number of SH3 domains, including the second and third Nck2 SH3 domains we studied previously (20), no regular secondary structure could be identified over the unique and long RT-loops, although they had extensive NOE connectivities with other parts of the molecules. As such, the loop residues still have highly limited backbone motions and, consequently, relatively high hNOE values. As seen in Fig. 6 *a*, except for the N-terminal residue Val^1, the C-terminal residue Asn^{56}, and Thr^{33} at the tip of the Src-loop, the wild-type SH3 domain at pH 6.5 uniformly had hNOE values >0.7, with about half of them >0.8, indicating that it adopted a well-folded structure at neutral pH. However, some residues at the tips of the RT-loop (Trp^7-Lys^{21}) and Src-loop (Asp^{29}-Arg^{36}) yielded slightly lower hNOE values, indicating that they had less limited backbone motions. By contrast, the hNOE values were significantly reduced for the wild-type at pH 2.0 and for 4AlaMut at pH 6.5, indicating that they acquired more freedom in backbone motions upon becoming partially folded. More specifically, for the wild-type at pH 2.0 and 4AlaMut at pH 6.5, the N-terminal and C-terminal residues had lower hNOE values, whereas the residues over two helices had higher hNOE values, with the majority >0.5 and some even >0.6, suggesting that the two helical fragments retained significantly limited backbone motions. Notably, the differences in hNOE values between helical and nonstructured regions are relatively small, implying that nonstructured regions are not random-coil with full freedom in backbone motions. Moreover, the $T2$ values also indicated similar patterns in which the residues located on two helical fragments had lower values than the N-terminal and C-terminal residues (Fig. 6 *c*). In 4AlaMut at pH 6.5, the $T2$ values for the residues over and close to the mutation region were larger than those of the corresponding residues in the wild-type at pH 2.0. On the other hand, for both the wild-type at pH 2.0 and 4AlaMut at pH 6.5, no uniform shortening of $T2$ values was observed, and the majority of their residues had $T2$ values larger than the corresponding residues of the wild-type at pH 6.5, implying that no significant aggregation existed in the two helical states under the conditions used for the NMR investigations here.

DISCUSSION

Previously, mainly based on studies of β-lactoglobulin, a nonhierarchical mechanism had been proposed for the fold-

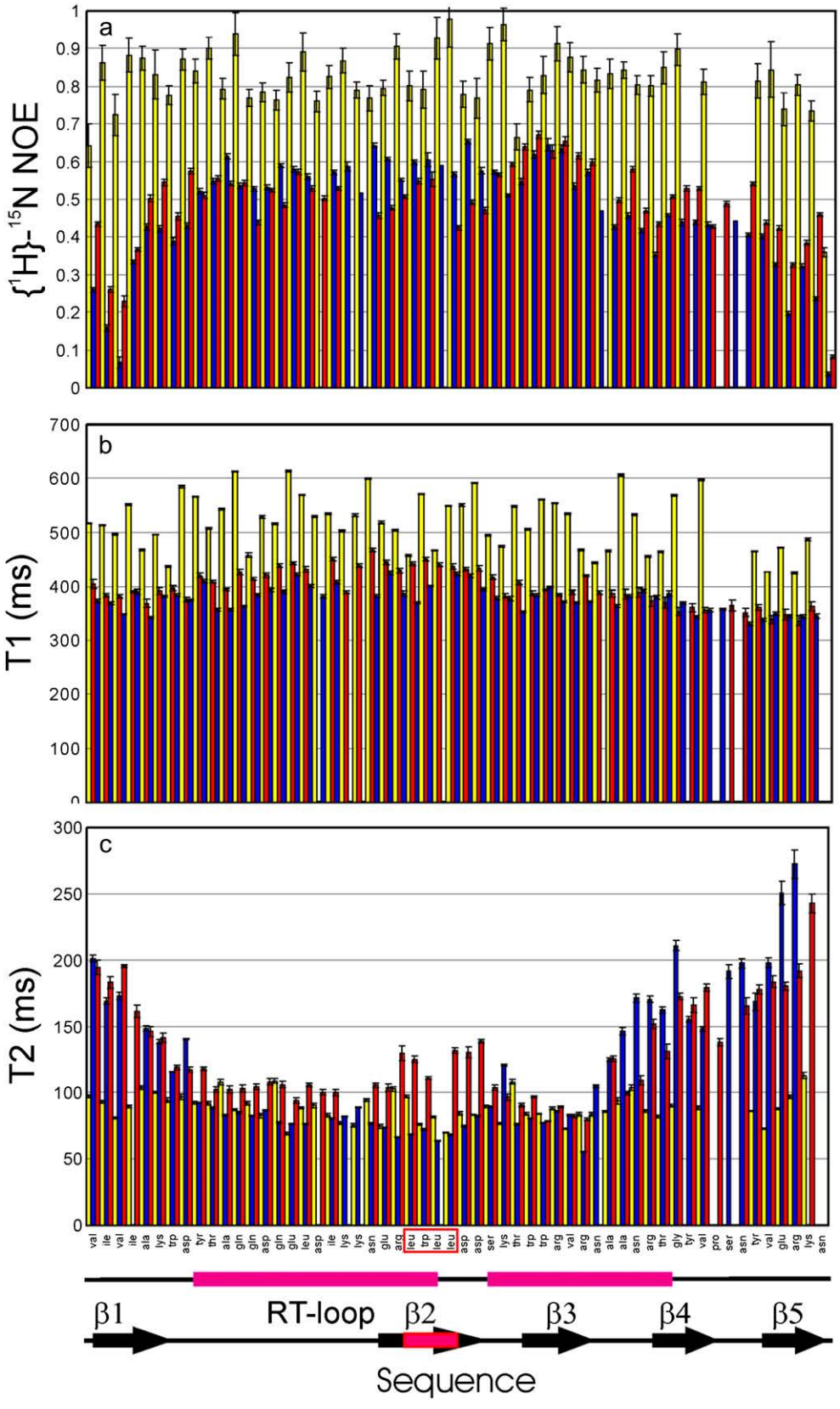

FIGURE 6 The ^{15}N NMR backbone relaxation data for wild-type at pH 6.5 (*yellow*), wild-type at pH 2.0 (*blue*), and 4AlaMut at pH 6.5 (*red*). (*a*) $\{^1H\}$-^{15}N steady-state NOE intensities. (*b*) ^{15}N *T*1 (longitudinal) relaxation times. (*c*) ^{15}N *T*2 (transverse) relaxation times. All NMR experiments used for deriving these data were recorded at 278 K on an 800-MHz Bruker Avance NMR spectrometer. Mutation region is enclosed in red, and secondary-structure fragments are indicated.

ing of β-proteins. In this mechanism, a nonnative helical intermediate is first formed from the collapse of a random coil, as driven by local interactions (11–17). The formation of this kind of nonnative helical intermediate is thought to be advantageous because it serves as a landmark to restrict the folding route, favoring further transformation into the native β-sheet structure, which is difficult to access kinetically (17). However, to date, the detailed conformational and dynamic properties for this sort of nonnative helical states are poorly documented.

In this study, the discovery that the wild-type form of the first hNck2 SH3 domain can be reversibly converted into a stable helical state at pH 2.0 offered an opportunity to investigate this sort of nonnative helical states at equilibrium by multidimensional NMR spectroscopy. The NMR parameters, such as chemical-shift deviations and NOE patterns, provided residue-specific evidence that the wild-type SH3 domain did switch into a highly populated two-helix state at pH 2.0. In this state, the first and last β-strands became highly unstructured, whereas the RT-loop, together with the second β-strand, was converted into the N-helix, and the third and fourth β-strands switched into the C-helix. The ^{15}N NMR backbone relaxation data further indicate that despite being partially folded, the residues on the two helices retain significantly limited backbone motions. Interestingly, a previous NMR study identified an ~50% populated helical conformation formed over N-terminal residues 4–14 of the α-spectrin SH3 domain (51). Here we speculate that the difference between the α-spectrin and Nck2 SH3 domains might mainly arise from their differential intrinsic properties for secondary-structure formation.

Regarding the nonhierarchical mechanism of β-protein folding, one of the most challenging tasks is to understand what sequence determinants specify the second transition, in which local and long-range interactions must interplay properly to disrupt the helical conformation and simultaneously facilitate the formation of the final β-structure. To address this issue, we performed a bioinformatics study, and succeeded in identifying a four-residue region on the second strand which has a high intrinsic propensity to form a β-strand. Indeed, mutation of these residues to Ala causes the SH3 domain to form a helical state even at pH 6.5. Moreover, extensive NMR studies demonstrated that the NMR conformational and dynamic properties of this mutant are highly similar to those of the wild-type SH3 domain at pH 2.0. This observation has two implications: 1), The formation of the nonnative helical state at pH 2.0 is unlikely to result from the specific effect imposed by the pH change. Instead, it is highly likely that acid unfolding nonspecifically destabilizes or disrupts the long-range interactions. Consequently, the intrinsic helix-formation propensity is released to drive the formation of the nonnative helical conformation observed at pH 2.0. 2), The four-residue region Leu^{25}-Trp^{26}-Leu^{27}-Leu^{28} appears to play no significant role in the first transition, i.e., the formation of the helical state, because its mutation still allows the correct formation of the helical conformation, highly similar to that of the wild-type at pH 2.0. However, these residues are essential for the second transition, to disrupt the helical conformation or to facilitate the formation of the final all-β SH3 fold by stabilizing the native β-structure. Moreover, our results highlight the complexity of the interplay between local and long-range interactions in specifying protein secondary and tertiary structures, consistent with a recent report that only a 12% sequence difference is able to switch a protein from a full-helical structure to an α/β fold with completely distinctive secondary and tertiary structures (52). Because the conformation of the two helical states determined here has no obvious similarity to the native all-β SH3 fold, it is challenging to speculate about the detailed mechanism underlying the second transition. However, examination of the native structure of the first hNck2 SH3 domain indicates that the four residues on the second β-strand have extensive long-range contacts with residues on other β-strands. A very interesting picture emerges with the inclusion of recently published results that mutating Ala^{45} to Gly would cause the Src SH3 domain to form a stable helical intermediate at pH 3.0 (25). As seen in Fig. 7, in both the hNck2 and Src SH3 domains, a tertiary hydrophobic core is formed among the side chains of the residues corresponding to the four residues identified in our Nck2 SH3 domain and Ala^{45} found in the Src SH3 domain. This strongly implies that the establishment of tertiary packing might represent a key step in the second transition. The tertiary packing core may serve to disrupt the helical conformation in the nonnative intermediate, and to direct the specific formation and assembly of the unique and complex SH3 fold. It will be of significant interest to map out which residues are involved in the formation of the tertiary core in the first hNck2 SH domain.

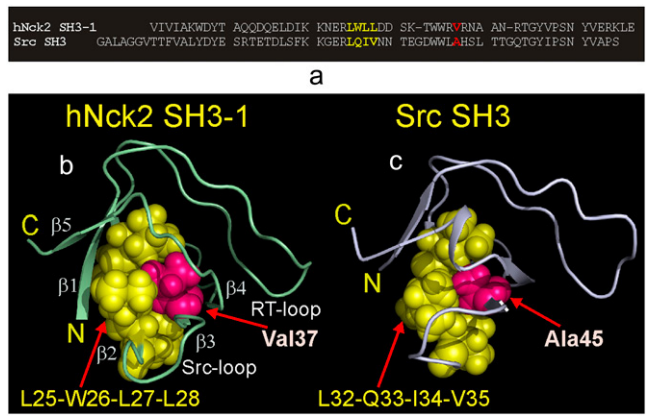

FIGURE 7 Tertiary packing cores in Src and first hNck2 SH3 domains. (*a*) Sequence alignment between Src and first hNck2 SH3 domains. (*b*) Tertiary packing cores in first hNck2 SH3 domain formed by residues Leu^{25}-Trp^{26}-Leu^{27}-Leu^{28} (*yellow*) identified in present study and Val^{37} (*red*), corresponding to Ala^{45} reported previously (25). (*c*) Tertiary packing cores in Src SH3 domain formed by residues Ala^{45} (*red*) and Leu^{32}-Gln^{33}-Ile^{34}-Val^{35} (*yellow*), corresponding to Leu^{25}-Trp^{26}-Leu^{27}-Leu^{28} in first Nck2 SH3 domain.

The current results may also have implications in understanding protein misfolding-related diseases, such as prion disease and Alzheimer's disease (53,54). For these diseases, the core mechanisms are associated with α-to-β conformation transitions. Previously it was found that for some protein fragments, one sequence is able to assume two distinctive secondary structures (hence termed the ''chameleon sequence'') (55). Given our observation that the uniquely defined all-β SH3 fold shared by more than 4000 sequences can highly populate a nonnative helical conformation, it is likely that the existence of the chameleon sequences has more profound connotations than previously recognized. This is in agreement with the recent conclusion that most proteins, if not all, can be converted to amyloid fibers under proper conditions, regardless of their native structures (53–55).

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

We thank Professor Franc Avbelj of the National Institute of Chemistry (Ljubljana, Slovenia) and Professor Robert L. Baldwin of the Stanford University School of Medicine (Stanford, CA) for providing the average chemical-shift tables.

This work was supported by Biomedical Research Council of Singapore Young Investigator Award R-154-000-217-305 to J.S.

REFERENCES

1. Kim, P. S., and R. L. Baldwin. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51:459–489.
2. Kim, P. S., and R. L. Baldwin. 1990. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59:631–660.
3. Kuwajima, K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins Struct. Funct. Genet.* 6:87–103.
4. Ptitsyn, O. B. 1995. Molten globule and protein folding. *Adv. Protein Chem.* 47:83–229.
5. Jennings, P. A., and P. E. Wright. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science.* 262:892–896.
6. Colón, W., and H. Roder. 1996. Kinetic intermediates in the formation of the cytochrome *c* molten globule. *Nat. Struct. Biol.* 3:1019–1025.
7. Song, J., P. Bai, L. Luo, and Z. Y. Peng. 1998. Contribution of individual residues to formation of the native-like tertiary topology in the α-lactalbumin molten globule. *J. Mol. Biol.* 280:167–174.
8. Song, J., N. Jamin, B. Gilquin, C. Vita, and A. Menez. 1999. A gradual disruption of tight side-chain packing: 2D 1H-NMR characterization of acidinduced unfolding of CHABII. *Nat. Struct. Biol.* 6:129–134.
9. Wei, Z., and J. Song. 2005. Molecular mechanism underlying the thermal stability and pH-induced unfolding of CHABII. *J. Mol. Biol.* 348:205–218.
10. Daggett, V., and A. Fersht. 2003. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* 4:497–502.
11. Kuwajima, K., H. Yamaya, and S. Sugai. 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.
12. Hamada, D., S. I. Segawa, and Y. Goto. 1996. Non-native α-helical intermediate in the refolding of β-lactoglobulin, a predominantly β-sheet protein. *Nat. Struct. Biol.* 3:868–873.
13. Arai, M., T. Ikura, G. V. Semisotnov, H. Kihara, Y. Amemiya, and K. Kuwajima. 1998. Kinetic refolding of β-lactoglobulin. Studies by synchrotron x-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy. *J. Mol. Biol.* 275:149–162.
14. Fujiwara, K., M. Arai, A. Shimizu, M. Ikeguchi, K. Kuwajima, and S. Sugai. 1999. Folding-unfolding equilibrium and kinetics of equine β-lactoglobulin: equivalence between the equilibrium molten globule state and a burst-phase folding intermediate. *Biochemistry.* 38:4455–4463.
15. Hamada, D., and Y. Goto. 1997. The equilibrium intermediate of β-lactoglobulin with non-native α-helical structure. *J. Mol. Biol.* 269:479–487.
16. Kuwata, K., R. Shastry, H. Cheng, M. Hoshino, C. A. Batt, Y. Goto, and H. Roder. 2001. Structural and kinetic characterization of early folding events in β-lactoglobulin. *Nat. Struct. Biol.* 8:151–155.
17. Chikenji, G., and M. Kikuchi. 2000. What is the role of non-native intermediates of β-lactoglobulin in protein folding? *Proc. Natl. Acad. Sci. USA.* 97:14273–14277.
18. Mayer, B. J. 2001. SH3 domains: complexity in moderation. *J. Cell Sci.* 114:1253–1263.
19. Musacchio, A. 2002. How SH3 domains recognize proline. *Adv. Protein Chem.* 61:211–268.
20. Liu, J., M. Li, X. Ran, J. Fan, and J. Song. 2006. Structural insight into the binding diversity between the human Nck2 SH3 domains and proline-rich proteins. *Biochemistry.* 45:7171–7184.
21. Li, W., J. Fan, and D. T. Woodley. 2001. Nck/Dock: an adapter between cell surface receptors and the actin cytoskeleton. *Oncogene.* 20:6403–6417.
22. Ran, X., and J. Song. 2005. Structural insight into the binding diversity between the Tyr-phosphorylated human ephrinBs and Nck2 SH2 domain. *J. Biol. Chem.* 280:19205–19212.
23. Park, S., K. Takeuchi, and G. Wagner. 2006. Solution structure of the first Src homology 3 domain of human Nck2. *J. Biomol. NMR.* 34:203–208.
24. Li, J. S., M. Shinjo, Y. Matsumura, M. Morita, D. Baker, M. Ikeguchi, and H. Kihara. 2007. An α-helical burst in Src SH3 folding pathway. *Biochemistry.* 46:5072–5082.
25. Li, J., Y. Matsumura, M. Shinjo, M. Kojima, and H. Kihara. 2007. A stable α-helix-rich intermediate is formed by a single mutation of the β-sheet protein, Src SH3, at pH3. *J. Mol. Biol.* 372:747–755.
26. Garnier, J., J. F. Gibrat, and B. Robson. 1996. GOR secondary structure prediction method version IV. *Methods Enzymol.* 266:540–553.
27. Rost, B., and C. Sander. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins.* 19:55–72.
28. Frishman, D., and P. Argos. 1996. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.* 9:133–142.
29. Levin, J. 1997. Exploring the limits of nearest neighbor secondary structure prediction. *Protein Eng.* 7:771–776.
30. Rohl, C. A., A. Chakrabartty, and R. L. Baldwin. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* 5:2623–2637.
31. Rohl, C. A., and R. L. Baldwin. 1998. Deciphering rules of helix stability in peptides. *Methods Enzymol.* 295:1–26.
32. Munoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* 245:275–296.
33. Munoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* 245:297–308.
34. Sattler, M., J. Schleucher, and C. Griesinger. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of

proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34:93–158.

35. Farrow, N. A., R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman Kay, and L. E. Kay. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry.* 33:5984–6003.
36. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6:277–293.
37. Johnson, B. A., and R. A. Blevins. 1994. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR.* 4:603–614.
38. Minor, D. L., Jr., and P. S. Kim. 1994. Measurement of the β-sheet-forming propensities of amino acids. *Nature.* 367:660–663.
39. Minor, D. L., Jr., and P. S. Kim. 1994. Context is a major determinant of β-sheet propensity. *Nature.* 371:264–267.
40. Venyaminov, S. Y., and J. T. Yang. 1996. Determination of protein secondary structure. *In* Circular Dichroism and the Conformational Analysis of Biomolecules. G. D. Fasman, editor. Plenum Press, New York. 69–107.
41. Wales, T. E., and J. R. Engen. 2006. Partial unfolding of diverse SH3 domains on a wide timescale. *J. Mol. Biol.* 357:1592–1604.
42. Sreerama, N., and R. W. Woody. 2000. Estimation of protein secondary structure from CD spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 282:252–260.
43. Merutka, G., H. J. Dyson, and P. E. Wright. 1995. ''Random coil'' ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR.* 5:14–24.
44. Schwarzinger, S., G. J. Kroon, T. R. Foss, P. E. Wright, and H. J. Dyson. 2000. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR.* 18:43–48.
45. Wishart, D. S., and A. M. Nip. 1998. Protein chemical shift analysis: a practical guide. *Biochem. Cell Biol.* 76:153–163.
46. Avbelj, F., D. Kocjan, and R. L. Baldwin. 2004. Protein chemical shifts arising from α-helices and β-sheets depend on solvent exposure. *Proc. Natl. Acad. Sci. USA.* 101:17394–17397.
47. Ramirez-Alvarado, M., F. J. Blanco, and L. Serrano. 1996. De novo design and structural analysis of a model β-hairpin peptide system. *Nat. Struct. Biol.* 3:604–612.
48. Jourdan, M., S. R. Griffiths-Jones, and M. S. Searle. 2000. Folding of a β-hairpin peptide derived from the N-terminus of ubiquitin. Conformational preferences of β-turn residues dictate non-native β-strand interactions. *Eur. J. Biochem.* 267:3539–3548.
49. Wagner, G., and K. Wuthrich. 1982. Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J. Mol. Biol.* 155:347–366.
50. Mandel, A. M., M. Akke, and A. G. Palmer III. 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246:144–163.
51. Blanco, F. J., L. Serrano, and J. D. Forman-Kay. 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.
52. Alexander, P. A., Y. He, Y. Chen, J. Orban, and P. N. Bryan. 2007. The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc. Natl. Acad. Sci. USA.* 104:11963–11968.
53. Chiti, F., and C. M. Dobson. 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–366.
54. Ventura, S., J. Zurdo, S. Narayanan, M. Parreno, R. Mangues, B. Reif, F. Chiti, E. Giannoni, C. M. Dobson, F. X. Aviles, and L. Serrano. 2004. Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc. Natl. Acad. Sci. USA.* 101:7258–7263.
55. Guo, J. T., J. W. Jaromczyk, and Y. Xu. 2007. Analysis of chameleon sequences and their implications in biological processes. *Proteins.* 67:548–558.