Article No. mb982072 *J. Mol. Biol.* (1998) **283**, 293–300

JMB

Prediction and Structural Characterization of an Independently Folding Substructure in the src SH3 Domain

Qian Yi, Chris Bystroff, Ponni Rajagopal, Rachel E. Klevit and David Baker*

Department of Biochemistry
University of Washington
Seattle, WA 98195, USA

**Corresponding author*

Previous studies of the conformations of peptides spanning the length of the α-spectrin SH3 domain suggested that SH3 domains lack independently folding substructures. Using a local structure prediction method based on the I-sites library of sequence-structure motifs, we identified a seven residue peptide in the src SH3 domain predicted to adopt a native-like structure, a type II β-turn bridging unpaired β-strands, that was not contained intact in any of the SH3 domain peptides studied earlier. NMR characterization confirmed that the isolated peptide, FKKGERL, adopts a structure similar to that adopted in the native protein: the NOE and $^{3}J_{NH_{\alpha}}$ coupling constant patterns were indicative of a type II β-turn, and NOEs between the Phe and the Leu side-chains suggest that they are juxtaposed as in the prediction and the native structure. These results support the idea that high-confidence I-sites predictions identify protein segments that are likely to form native-like structures early in folding.

Keywords: SH3 domain; folding initiation site; protein folding; local structure; β-turn

Introduction

There has been considerable discussion about the role of local interactions in protein folding (Abkevich *et al*., 1995; Avbelj & Moult, 1995; Doyle *et al*., 1997; Fersht, 1995; Muñoz & Serrano, 1996; Unger & Moult, 1996). Direct experimental studies are hampered because early events of protein folding are likely to take place within the microsecond time-scale. To investigate the role of β-turns, peptide fragments derived from proteins have been studied as models for early events in protein folding (Dyson *et al*., 1988, 1992). Those studies have shown that sequences that form turns in proteins can also form reasonably stable turns in short peptides in aqueous solution. However, because non-local interactions play an important role in stablizing structure in both the native state and denatured state (Wang & Shortle, 1997), many peptides derived from proteins do not adopt well-defined structure in isolation. In particular, a recent study of five peptides derived from the α-spectrin SH3 domain failed to detect any persistent structure and it was concluded that folding initiation sites do not play a role in the folding of this all-β protein (Viguera *et al*., 1996).

We have recently developed a method for local protein structure prediction based on a library (I-sites) of 7 to 19 residue sequence patterns that strongly correlate with local protein structural features (Bystroff & Baker, 1997, 1998). The sequence segments matching a particular pattern almost always adopt the same conformation in a wide range of protein structures, suggesting that local interactions within the segment are strong enough to override the differences in non-local interactions. Inspection of the sequence-structure motifs in most cases readily reveals the interactions that stabilize the observed structure. One of the novel motifs is a "diverging type II β-turn" stabilized by a side-chain-to-backbone hydrogen bond and a pair of inwardly turned hydrophobic residues bracketing the turn (Bystroff & Baker, 1998). The turn is

Abbreviations used: 1D, one-dimensional; ppb, parts per billion; ppm, parts per million; NOE, nuclear Overhauser effect; TOCSY, total correlated spectroscopy; ROESY, rotating frame Overhauser effect spectroscopy; TSP, 3-(Trimethylesilyl)propionic-2,2,3,3-d_4 acid; MALDI-MS, matrix assisted laser desorption ionization mass spectrometry.

E-mail address of the corresponding author: baker@ben.bchem.washington.edu

0022–2836/98/410293–08 $30.00/0

referred to as diverging because the two strands connected by it do not form backbone hydrogen bonds.

We have proposed that, since the interactions within the sequence segments override non-local interactions, peptide segments that closely match one of the sequence patterns are likely to adopt structure in isolation and potentially serve as folding initiation sites. Because the SH3 domain had been proposed to lack such peptide segments, we were curious about I-sites predictions of local structure for this protein family. Interestingly, the peptide segment predicted to be the most likely to have structure in isolation was a diverging type II β-turn that was not contained intact in any of the α-spectrin SH3 peptides studied previously (Viguera *et al*., 1996). In this study, we demonstrate that the sequence FKKGERL, derived from the diverging type II turn in the src SH3 domain, adopts that conformation in isolation.

Results

Local structure prediction

Figure 1(a) shows the location of all correct (red) and incorrect (blue) fragment structure predictions (see Materials and Methods) for the src SH3 domain. The two highest-confidence predictions occurred around the position of the type II β-turn at residues 27 to 30 (shaded region in Figure 1(b)); both correctly predict a type II turn, an inwardly turned glutamate side-chain, and the juxtaposition of the Phe26 and Leu32 side-chains. We selected the peptide FKKGERL (residues 26 to 32) for structural studies, omitting three residues of the β-strand. The high-confidence prediction is a consequence of the similarity between the sequence pattern for the I-sites diverging turn motif (Figure 2(a)) and the SH3 multiple sequence alignment at residues 26 to 32 (Figure 2(c)). As expected, given the strong similarity of the sequence patterns, the predicted structure (Figure 2(b)) is very similar to that in the crystal structure of the src SH3 domain (Figure 2(d)).

NMR study of the peptide conformation

To avoid artifacts due to non-native electrostatic interactions between the N and C termini, we studied the structure of two forms of the peptide; one with free N and C termini, the other with an acetylated N terminus and an amidated C terminus. The results were virtually identical for the two peptides; for brevity, we present only the data for the blocked peptide.

Backbone conformation

The proton NMR spectrum of the peptide was completely assigned using a 50 ms TOCSY experiment and confirmed by a 250 ms ROESY experiment. The chemical shifts for all proton resonances

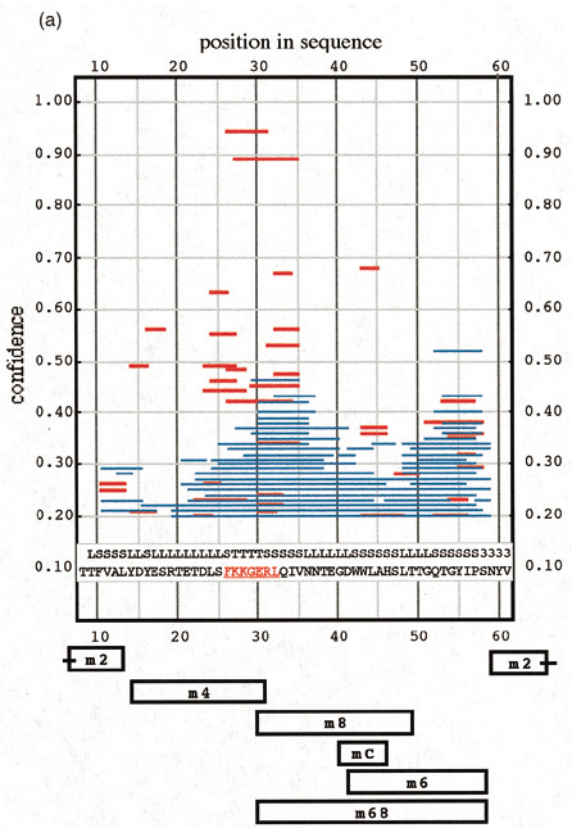

Figure 1. (a) I-sites fragment predictions for the SH3 domain sequence family. Red bars indicate correct predictions according to the crystal structure (1FMK); blue bars are incorrect predictions. In the text box is the sequence and secondary structure of the src SH3 domain; (S, strand; T, turn; L, loop; 3, 3_{10} helix). The peptides of α-spectrin SH3 domain studied by Viguera *et al*. (1996) are indicated by the rectangular bars below the graph. (b) Backbone trace of the SH3 domain (1FMK) showing the location of the diverging turn (shaded).

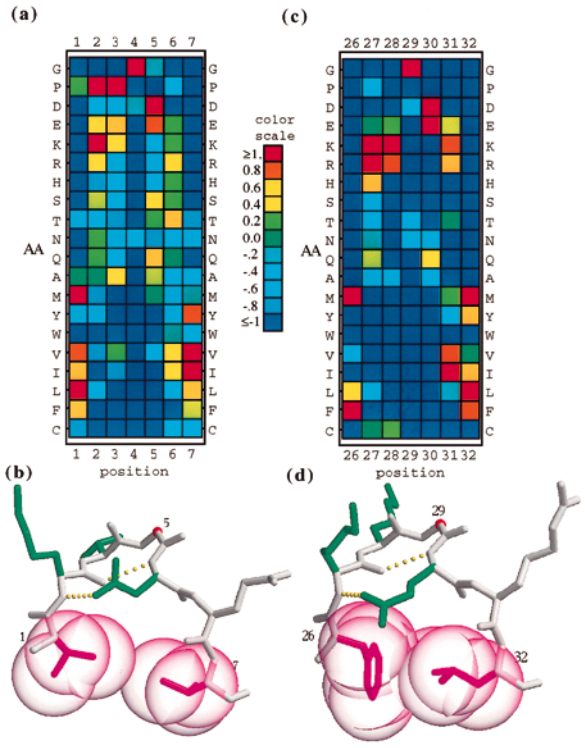

Figure 2. (a) Sequence profile and (b) paradigm structure (1EFT residues 247 to 253) for the diverging turn motif from the I-sites library (Bystroff & Baker, 1998). (c) Sequence profile for the diverging turn segment in the SH3 domain sequence family, and (d) the native structure of this portion of the src SH3 domain (1FMK residues 102 to 108; Xu *et al.*, 1997). Numbering in (c) and (d) is consistent with our previous studies (Riddle *et al.*, 1997). The colors in the profile tables represent the log likelihood ratio, log $[P_{ij}/P_i]$, where P_{ij} is the frequency of amino acid i at position j in the I-sites sequence pattern (a) or SH3 domain multiple sequence alignment (c), and P_i is the average frequency of amino acid i in the proteins. The color scale is shown in the central panel; favored amino acids are in red and disfavored in blue. Conserved polar residues in the sequence profiles are shown in green; conserved non-polar residues in purple. Solvent-accessible surfaces are drawn around the non-polar side-chains. The Figure was made using Raster3D (Merritt & Murphy, 1994) and Mathematica (Wolfram Research, Inc).

are summarized in Table 1. The NOE pattern observed in ROESY spectra, the three-bond NH^{α} coupling constants, and the temperature coefficients of the amide protons are summarized in Figure 3. As described in the following paragraphs, all the NMR parameters indicate that a type II β-turn conformation is significantly populated.

β-Turns yield characteristic patterns of sequential and medium-range NOEs and $^3J_{NH_\alpha}$ coupling constants (Wüthrich, 1986). NOE patterns indicative of β-turns include a d_{NN} NOE between residues 3 and 4 and a $d_{\alpha N}(i, i+2)$ NOE between residues 2 and 4 in the turn. Type I and II turns are distinguished by the relative strengths of the $d_{\alpha N}(2,3)$ NOE (stronger for type II) and the $d_{NN}(2,3)$ NOE (stronger for type I). A strong $d_{NN}(i, i+1)$ NOE between Gly29 and Glu30, and a less strong $d_{\alpha N}(i, i+2)$ NOE between Lys28 and Glu30 in the peptide were observed in ROESY spectra (Figure 4(a)). The presence of this characteristic NOE pattern suggests a high preference for a β-turn conformation for the Lys-Lys-Gly-Glu segment in the peptide. The pattern of a strong $d_{\alpha N}$ and a weak d_{NN} NOE between Lys28 and Gly29, along with a very weak $d_{\beta N}$ NOE between Lys28 and Glu30 (observed at lower thresholds, but not shown in Figure 4(a)), is most consistent with a type II β-turn (Campbell *et al.*, 1995; Wüthrich, 1986). β-Turns are characterized also by a small value (~5 Hz) of $^3J_{NH_\alpha}$ for residue 2 in the turn (Wüthrich, 1986). The observed $^3J_{NH_\alpha}$ coupling constant for Lys28 (5.0 Hz, Figure 3), compared to the $^3J_{NH_\alpha}$ value of 6.6 Hz for Lys in the random coil conformation (Smith *et al.*, 1996), is consistent with its position as residue 2 in a turn conformation.

In most β-turns there is a backbone-backbone hydrogen-bond between the carbonyl oxygen atom of residue 1 and the NH group of residue 4. This hydrogen bond often leads to a small temperature coefficient $(0 < -\Delta\delta/\Delta T < 5$ ppb $K^{-1})$ for the amide proton of residue 4 (Rose *et al.*, 1985) . The amide proton of Glu30 has a temperature coefficient of -4.6 ppb K^{-1} (Figure 3) (compared to $6 < -\Delta\delta/\Delta T < 10$ ppb K^{-1} for amide groups in a random coil conformation), suggesting a role as a hydrogen-bond donor, as expected for residue 4 in a turn conformation. This, combined with the evidence for a type II turn conformation, would implicate the carbonyl oxygen atom of Lys27 as the likely H-bond acceptor.

Table 1. Proton assignments in 50 mM sodium phosphate at pH 6.0 and 12°C

Residue	NH	$C^{\alpha}H$	$C^{\beta}H$	Others
Phe26	8.34	4.54	3.09, 3.01	δ7.26, ε7.36, ζ7.32
Lys27	8.45	4.30	1.78, 1.68	γ1.38, δ1.65, ε2.99
Lys28	8.48	4.20	1.80, 1.73	γ1.48; 1.45, δ1.71, ε3.03
Gly29	8.65	3.93, 3.99		
Glu30	8.04	4.31	1.95, 2.04	γ2.09; 2.25
Arg31	8.55	4.31	1.84, 1.78	γ1.61, δ3.19, $N^{\varepsilon}H$ 7.25
Leu32	8.41	4.31	1.67, 1.60	γ1.62, δ0.93; 0.87

TSP was used as a reference for chemical shifts.

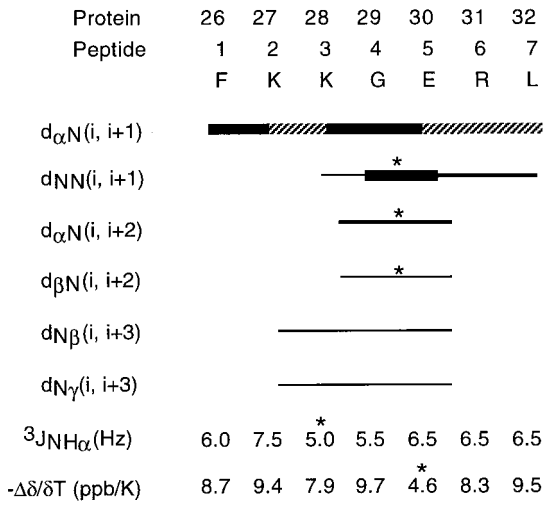

Figure 3. Observed NOE patterns, values of the $^3J_{NH_\alpha}$ coupling constants, and the temperature coefficients for the peptide FKKGERL. The widths of the lines represent the NOE intensities. The broken lines represent some of the $d_{\alpha N}(i-1, i)$ NOEs that overlap with $d_{\alpha N}(i, i)$ NOEs. The results indictive of the type II β-turn conformation are indicated with asterisks (*).

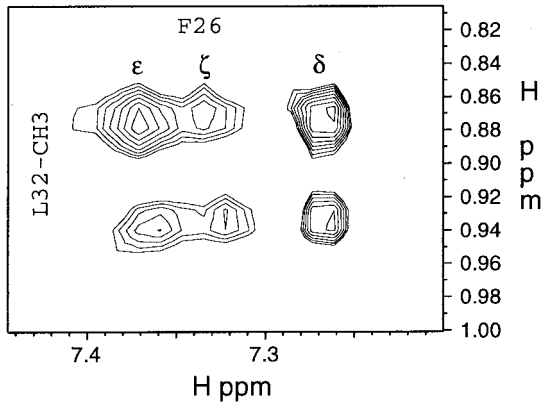

Figure 5. Portion of the ROESY spectrum displaying the NOEs between the aromatic protons of Phe26 and the methyl groups of Leu32 at 12°C and pH 6.0.

Side-chain interactions that stabilize the β-turn

In addition to the backbone hydrogen bond (Lys27 CO-Glu30 NH) interaction described above, there are two other interactions observed in the folded protein that may contribute to the stability of the isolated β-turn (Figure 2(d)). The first is a hydrophobic interaction between the aromatic ring of Phe26 and the methyl groups of Leu32. NOE cross-peaks between the aromatic protons of Phe26 and the methyl groups of Leu32 are clearly visible in the ROESY spectrum at 12°C (Figure 5), indicating that conformations are populated in which the aromatic ring and the methyl groups are in relatively close contact (<5 Å). The side-chain NOE between Phe26 and Leu32 is observed at 12°C but

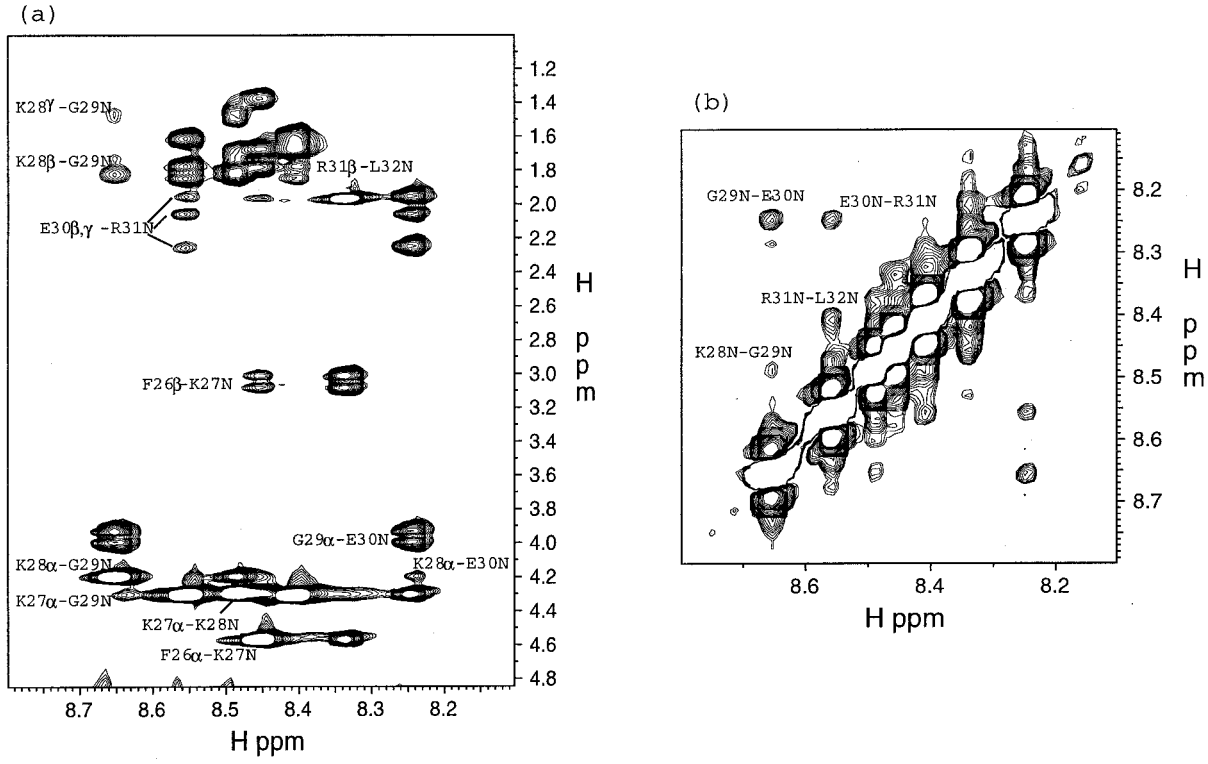

Figure 4. ROESY spectrum at pH 6.0 and 12°C. (a) the C(α, β)H - NH region, (b) the NH - NH region. A 250 ms mixing time was used. Only the inter-residue NOEs are labelled in the spectra.

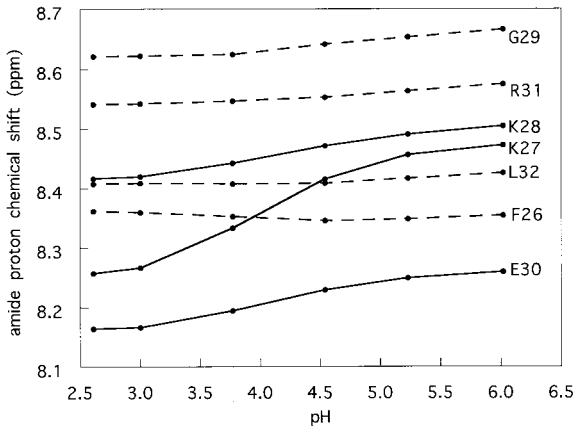

Figure 6. Dependence of the amide proton chemical shifts on pH. The amide protons of K27, K28 and E30 (continuous curves)are sensitive to pH, while the amide groups of F26, G29 and L32 (broken curves) are not.

not at 24°C (data not shown). These results suggest that the two hydrophobic side-chains are juxtaposed as in the prediction and in the native structure.

The second native interaction is a hydrogen bond between a side-chain carboxylate oxygen atom of Glu30 and the backbone amide proton of Lys27. We reasoned that such an interaction would be dependent on the ionization state of the Glu30 side-chain, the only ionizable group in the peptide with a pK_a between 2.5 and 6.5, and carried out a pH titration to investigate this possibility. The amide proton of Lys27 undergoes the largest chemical shift change of any amide over a pH range of 2.5 to 6.5. When the Glu30 side-chain is deprotonated at pH 6.0, the amide proton of Lys27 in the peptide is shifted downfield by ~0.22 ppm relative to pH 2.6 (Figure 6). This shift is consistent with an interaction with the Glu30 side-chain, as hydrogen-bonding will cause deshielding of the amide proton, which generally results in a downfield shift for the amide proton. The temperature coefficient of Lys27 amide proton is greater than the expected value for an amide proton involved in hydrogen-bonding in proteins (-9.4 ppb K^{-1} *versus* -5.0 ppb K^{-1}) but, since an increase in temperature is likely to increase the mobility of the Glu30 side-chain significantly, the temperature coefficient is probably not an accurate indicator of side-chain to main-chain hydrogen bond formation in the peptide. Weak NOEs between the β and γ protons of the Glu30 side-chain and the amide proton of Lys27 were observed at low threshholds in ROESY spectra (Figure 3). These results further suggest that the side-chain of the Glu30 is pointed toward the Lys27 amide proton, consistent with a hydrogen bond between the Lys27 amide proton and the side-chain carboxyl oxygen atom of Glu30.

Discussion

The structure of fragments of the src-SH3 domain was predicted using the I-sites library. The peptide fragment with the highest confidence prediction, FKKGERL, was synthesized and its structure characterized by proton NMR. The NMR parameters support the prediction that the peptide adopts a native-like diverging type II β-turn (Figure 2(b)). This is the first short peptide from an SH3 domain that has been demonstrated to adopt a native-like conformation in isolation.

I-sites predictions and the conformational preferences of isolated peptides

The diverging turn appears to be the lowest free energy conformation of the short peptide studied here. There are three factors that favor the diverging turn conformation for the peptide FKKGERL (Figure 2(b)): the hydrophobic interaction between the Phe at position 1 and the Leu at position 7, the Gly at position 4 allowing a positive phi angle, and the Glu at position 5 that forms a hydrogen bond with the backbone amide group at position 2. The pattern of sequence conservation in the I-sites profile shows each of these features, but it shows also features that can be explained only by negative design. Polar side-chains are conserved in positions 3 to 5 (Figure 2(a)); this prevents the formation of a stable amphipathic α-helix, since the latter requires conserved non-polar side-chains separated by three or four residues. Similarly, polar side-chains in positions 3 and 5 destabilize the amphipathic β-strand conformation, which prefers non-polar residues in those positions. Position 3 in the profile is always polar, but is never observed to be Asp or Asn. This may be negative design against the type I β-turn, which prefers Asp or Asn followed by Gly.

The lack of stable structure in the SH3 peptides studied by Serrano and colleagues (Viguera *et al.*, 1996) is consistent with the I-sites predictions. As illustrated in Figure 1(a), there is no high-confidence prediction spanning the five peptides that they studied. Two of the peptides contain portions of the diverging turn; however, neither contains the whole segment (the sequence MKKGDIL in the α-spectrin SH3 domain).

The confidence of a fragment prediction may be related to the fragment's free energy of folding in isolation. The confidence of I-sites predictions is defined as the fraction of peptide segments with a certain similarity score that have the predicted structure (Bystroff & Baker, 1998). For example, all seven residue peptide segments in the protein database were scored against the sequence profile for the diverging turn motif (Figure 2(a)), and 94% of all segments with a similarity score between 144 and 151 were found to have the diverging turn structure; therefore, a new sequence segment with a score of 147 has an estimated 94% probability of being a diverging turn. Because the segments in

the database from which the I-sites library was derived all differ in their respective global structures, the strong structural propensities must be due to internal contacts conserved within I-sites clusters. Clearly, the relation of confidence to equilibrium concentration, and through this to free energy, is a crude one. Nonetheless, the evidence suggests that the relation holds qualitatively, at least for sequence segments with high confidence scores.

The studies by Dyson *et al.* (1992) on short peptides comprising the entire length of the β-sandwich protein plastocyanin showed that only a small number of peptides had any tendency to form structure in isolation. Interestingly, the peptide with the strongest structural tendencies was a diverging β-turn similar to that studied here. To further explore the relationship between I-sites predictions and peptide conformational preferences, I-sites predictions were made for the plastocyanin sequence family. The three regions found previously to have some native structural tendencies in isolation were all contained within correct I-sites structure predictions with a confidence of 0.60 or greater.

We present several more I-sites predictions (Table 2) to be confirmed or refuted by future experimental evidence. These short sequences were chosen mostly from small proteins, some of which are presently under intensive experimental study. They are predicted to adopt a significant amount of native structure in isolation. Structure-blind I-sites predictions can be made automatically *via* http://ganesh.bchem.washington.edu/ ~ bystroff/Isites/.

Role of diverging turn in protein folding

It is interesting to compare our results with recent structural information obtained on unfolded states of the drkN SH3 domain under both folding conditions (U_{exch}) and denaturing conditions (U_{Gnd}; Zhang & Forman-Kay, 1997). The most marked differences between the U_{exch} and U_{Gnd} states are located in the segment $_{19}$FRKTQILKIL$_{28}$, which corresponds to $_{26}$FKKGERLQIV$_{35}$ in the src SH3 domain (the diverging turn segment is underlined). A turn conformation was populated around $_{19}$FRKT$_{22}$ in U_{exch}, but not in U_{gnd} , and residues in $_{23}$QILKIL$_{28}$ exhibited severe line-broadening in the U_{exch} state that disappeared upon addition of denaturant, suggesting that this portion of the chain adopts a structure that is in intermediate exchange with other conformational substates. By analogy to our results, the conformational substates populated by this portion of the drk U_{exch} may include a similar type of diverging turn conformation, which may be less stable in drk than in src because of the replacement of the glycine residue by threonine (Figure 2(a)).

The characteristic features of the diverging turn sequence motif are highly conserved in all known SH3 domains (Koyama *et al.*, 1993). Each of the other hairpin turns in the molecule shows evolutionary variability in both sequence and length. This portion of the structure is less variable among the three-dimensional structures of SH3 domains solved to date than any other segments outside of the hydrophobic core (Guruprasad *et al.*, 1995). The strong sequence and structural conservation within the diverging turn along with the drk denatured state and the src peptide studies may indicate an important role for the turn in the folding process.

It is interesting that the turn in the SH3 domain with the strongest propensity to adopt structure in isolation is not a tight β-hairpin; instead, the turn connects strands that do not share backbone hydrogen bonds (the earlier peptide studies on α-spectrin SH3 focused on the β-hairpins). The SH3 domain may be viewed as two orthogonally packed β-sheets, where the diverging β-turn is one of the two transitions between the sheets in the protein. Formation of the diverging β-turn conformation early in folding could play an important role in establishing the topology of the protein by preventing inappropriate formation of a β-hairpin and promoting proper packing of hydrophobic side-chains between the diverging strands. The

Table 2. Blind predictions of folding initiation sites

PDB code	Seq. numbers	Sequence	I-sites motif/native structure	*Cf*[a]
1aaj	37-45	VKVGDTVTWI	Diverging turn	1.00
1aaj	72-80	MKKEQAYSL	Diverging turn	0.85
1edt	77-87	RPLQQQGIKVL	Helix C-cap, type 1	1.00
1edt	50-57	YDTGTKTA	Ser beta-hairpin	0.80
1htp	106-112	TSPDELE	Helix N-cap	1.00
1htp	115-121	LGAKEYTKFC	Helix N-cap	1.00
1ifc	14-25	YEKFMEKMGINV	Helix C-cap, type 2	1.00
3aah	393A-401A	YDPESRTLY	Ser-hairpin	1.00
1bgl	576A-585A	SLIKYDENGNPW	Type I hairpin	1.00
1nhp	63-74	GEKMESRGVNVF	Helix C-cap, type 2	1.00
1trk	316A-326A	FSEYQKKFPEL	Pro helix C-cap	0.90

Columns 1 and 2 give the PDB code and residue names, and column 3 gives the sequences of segments predicted to adopt native structure in isolation. The I-sites sequence-structure motif is listed in column 4, and column 5 is the confidence of the prediction. These fragments were selected at random from a large set of predictions. A prediction server is available at the web site http://ganesh.bchem.washington.edu/ ~ bystroff/Isites/.

observation of a structured diverging turn peptide in plastocyanin suggests that such a role may be a common feature of the folding of β-sheet proteins. Our results show that the diverging turn in src-SH3 is stable in isolation and, because all of the interactions are local, it undoubtedly forms very rapidly. This is consistent with kinetic studies of src SH3 mutants (Grantcharova *et al.*, 1998), which suggest that the diverging turn and the following strand come together with the distal loop hairpin in the folding transition state.

An attractive feature of the protein folding problem is that it is amenable to both computational and experimental approaches. The work described here illustrates that a combination of these approaches can provide significantly more insight than either one alone. *Ab initio* structure prediction methods even in their current imperfect state can generate hypotheses to guide experimental studies of the folding process.

Materials and Methods

Prediction of peptide conformation

A sequence profile (Gribskov *et al.*, 1990) was constructed from an alignment of SH3 domain sequences in the SWISPROT database, aligned initially *via* the PHD server (Rost *et al.*, 1994) then modified to agree with earlier structural alignments (Feng *et al.*, 1995; Guruprasad *et al.*, 1995). All subfragments of the profile were scored against all motifs in the library as described elsewhere (Bystroff & Baker, 1998). Scores were translated into confidence values using the results of cross-validation studies on a large non-redundant database of proteins of known structure; the confidence of a prediction is simply the probability that the prediction is correct.

NMR examination of the peptide conformation

The sequence Phe-Lys-Lys-Gly-Glu-Arg-Leu was synthesized and purified by Research Genetics Co. (Huntsville, AL), and the molecular mass was confirmed by MALDI-MS. Two forms of the peptide, one with free amino and carboxyl groups at the N and C termini, the other with an acetylated N terminus and an amidated C terminus, were investigated by NMR. NMR samples were prepared by dissolving ~10 mg of peptide in 0.45 ml of 50 mM sodium phosphate buffer (90% H_2O/10% ^2H_2O). The pH of the samples were adjusted using diluted NaOH or HCl solutions. TSP was added to the NMR samples to a final concentration of 0.5 mM for chemical shift referencing.

NMR spectra were acquired on a Bruker DMX500 spectrometer at 12°C unless otherwise specified. ^{1}H-TOCSY spectra (Bax & Davis, 1985a) were collected using spectral widths of 6250 Hz in both dimensions and 1024 × 600 complex points. ^{1}H-ROESY spectra (Bax & Davis, 1985b) were collected with a spin-lock field strength of 8.62 kHz, a spectral width of 6250 Hz in both dimensions and 1024 × 600 complex points. The mixing time for the TOCSY and ROESY experiments were 50 ms and 250 ms, respectively. Data were apodized with a squared sine bell in both dimensions, and zero-filled to give 1 K × 1 K complex spectra. Water supression was achieved with the watergate pulse sequence (Sklenar *et al.*, 1993). The recycle delay was 2.2 s for all the experiments. All the data processing was performed on an SGI workstation using the program NMRpipe (Delaglio *et al.*, 1995).

The $^3J_{NH_\alpha}$ coupling constants were obtained directly from the resolved amide proton resonances in the 1D spectrum collected with a digital resolution of 0.43 Hz/point. The temperature coefficients of the amide protons were obtained from linear fits of the chemical shift data from 1D spectra acquired at 9, 12, 15, 18, 21, 24, 27, 30, 33 and 36°C.

Acknowledgements

We thank David Shortle, Patricia Campbell, Kevin Plaxco and Kim Simons for helpful comments on the manuscript. This work was supported by young investigator awards to D.B. from the NSF and Packard foundation.

References

Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1995). Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. *J. Mol. Biol.* **252**, 460–471.

Avbelj, F. & Moult, J. (1995). Determination of the conformation of folding initiation sites in proteins by computer simulation. *Proteins: Struct. Funct. Genet.* **23**, 129–141.

Bax, A. & Davis, D. G. (1985a). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360.

Bax, A. & Davis, D. G. (1985b). Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson.* **63**, 207–213.

Bystroff, C. & Baker, D. (1997). Blind ab initio local structure predictions using a library of sequence-structure motifs. *Proteins: Struct. Funct. Genet.* **(Suppl. 1)**, 167–171.

Bystroff, C. & Baker, D. (1998). Prediction of local structure in proteins using a library of sequence-structure motifs. *J. Mol. Biol.* **281**, 565–577.

Campbell, A. P., McInnes, C., Hodges, R. S. & Sykes, B. D. (1995). Comparison of NMR solution structures of the receptor binding domains of Pseudomonas aeruginosa pili strains PAO, KB7, and PAK: implications for receptor binding and synthetic vaccine design. *Biochemistry*, **34**, 16255–16268.

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

Doyle, R., Simons, K., Qian, H. & Baker, D. (1997). Local interactions and the optimization of protein folding. *Proteins: Struct. Funct. Genet.* **29**, 282–291.

Dyson, H. J., Rance, M., Houghten, R. H., Lerner, R. A. & Wright, P. E. (1988). Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* **201**, 161–200.

Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H. C., Lerner, R. A. & Wright, P. E. (1992). Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin. *J. Mol. Biol.* **226**, 819–835.

Feng, S., Kasahara, C., Rickles, R. J. & Schreiber, S. L. (1995). Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. *Proc. Natl Acad. Sci. USA,* **92**, 12408–12415.

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

Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nature Struct. Biol.* **5**, 714–720.

Gribskov, M., Luthy, R. & Eisenberg, D. (1990). Profile analysis. *Methods Enzymol.* **183**, 146–159.

Guruprasad, L., Dhanaraj, V., Timm, D., Blundell, T. L., Gout, I. & Waterfield, M. D. (1995). The crystal structure of the N-terminal SH3 domain of Grb 2. *J. Mol. Biol.* **248**, 856–866.

Koyama, S., Yu, H., Dalgarno, D. C., Shin, T. B., Zydowsky, L. D. & Schreiber, S. L. (1993). Structure of the PI3 K SH3 domain and analysis of the SH3 family. *Cell,* **72**, 945–952.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. Sect. D,* **6**, 869–873.

Muñoz, V. & Serrano, L. (1996). Local versus nonlocal interactions in protein folding and stability – an experimentalist's point of view. *Folding Des.* **1**, R71–R77.

Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q. & Baker, D. (1997). Functional rapidly folding proteins from simplified amino acid sequences. *Nature Struct. Biol.* **4**, 805–809.

Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985). Turns in peptides and proteins. *Advan. Protein Chem.* **37**, 1–106.

Rost, B., Sander, C. & Schneider, R. (1994). PHD–an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10**, 53–60.

Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. (1993). Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. Ser. A,* **102**, 241–245.

Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M. & Dobson, C. M. (1996). Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255**, 494–506.

Unger, R. & Moult, J. (1996). Local interactions dominate folding in a simple protein model. *J. Mol. Biol.* **259**, 988–994.

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

Wang, Y. & Shortle, D. (1997). Residual helical and turn structure in the denatured state of staphylococcal nuclease: analysis of peptide fragments. *Folding Des.* **2**, 93–100.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.

Xu, W., Harrison, S. C. & Eck, M. J. (1997). Three dimensional structure of the tyrosine kinase C-src. *Nature,* **385**, 595–602.

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

Edited by P. E. Wright

(Received 10 April 1998; received in revised form 2 July 1998; accepted 7 July 1998)