Designing a 20-residue protein

Jonathan W. Neidigh, R. Matthew Fesinmeyer and Niels H. Andersen

Department of Chemistry, University of Washington, Seattle, Washington 98195, USA

Published online: 29 April 2002, DOI: 10.1038/nsb798

Truncation and mutation of a poorly folded 39 residue peptide has produced 20 residue constructs that are >95% folded in water at physiological pH. These constructs optimize a novel fold, designated as the 'Trp cage' motif, and are significantly more stable than any other miniprotein reported to date. Folding is cooperative and hydrophobically driven by the encapsulation of a Trp side chain in a sheath of Pro rings. As the smallest protein like construct, Trp cage miniproteins should provide a testing ground for both experimental studies and computational simulations of protein folding and unfolding pathways. Pro Trp interactions may be a particularly effective strategy for the *a priori* design of self folding peptides.

The *a priori* design of proteins and the determination of the minimum requirements for the formation of protein like structures are currently the subject of active research, with several notable achievements reported in the past few years[1–5]. Excluding systems with disulfide or metal chelation crosslinks, the smallest natural domains that fold autonomously have 32–40 residues[6–9]. The design of stable protein like structures with a smaller number of residues requires both an exquisite optimization of hydrophobic packing and a strategy to decrease the backbone configurational entropy of the unfolded state.

There have been successes in the minimization of known folds and the design of stable miniproteins. Dahiyat and Mayo[1] have reported a 28 residue zinc finger mimic that displays a stable fold and some degree of cooperativity in its melting behavior as monitored by CD. The smallest natural proteins are WW domains, which have an antiparallel three stranded β sheet and two conserved Trp residues. The Pin WW domain can be truncated to 28 residues and retain two state folding behavior[9]. The Serrano group designed Betanova[10], a 20 residue β sheet that is ~12% folded in water at 10 °C (ref. 11). Well folded three stranded β sheets of 20–29 residues in length have been designed[4,12], but these require D amino acids to improve turn propensities. The goal of our fold optimization program is the production of a 20 residue, self folding peptide using only the standard set of L amino acids. Further, we want the hydrophobic effect[13,14], rather than secondary structure stability, to drive structure formation to impart protein like folding behavior.

Although there is no universally accepted definition for a 'self folding domain'[7], the following features are useful diagnostics: (i) multiple secondary structure elements in close contact; (ii) notable chemical shift dispersion that predominantly reflects tertiary interactions; (iii) side chain–side chain packing interactions that produce well defined χ^1 and χ^2 values; (iv) greater backbone amide exchange protection than that provided by secondary structure; and (v) some degree of folding cooperativity and resistance to thermal unfolding. Here we report a 20 residue construct (Fig. 1*b*) that displays these characteristics and the fold minimization strategy that provided this miniprotein.

Mutational optimization of the 'Trp cage'

The NMR structure[15] of the predominantly helical, 39 residue peptide exendin 4 (EX4) from Gila monster saliva served as our starting point. EX4 appears to be poorly folded in water as it aggregates at NMR concentrations but is stable and folded (Fig. 1*a*) in aqueous fluoroalcohol media. The feature that was most striking about the XFXXWXXXXGPXXXXPPPX (where X is any amino acid) sequence is the close association of a Gly and three Pro residues (shown in bold) with the two aromatic side chains. We have designated this fold as the 'Trp cage'. Sequence remote hydrogens located above and below the indole ring of Trp 25 display large chemical shift deviations (CSDs) due to ring current shielding effects. The following CSDs were observed for EX4 in 30% trifluoroethanol (TFE): −3.00 and −1.44 (Gly 30 α2 and α3, respectively), −0.67 (Pro 31 δ3), −1.82 (Pro 37 α), −1.52 (Pro 37 β3) and −0.57 p.p.m. (Pro 38 δ2/δ3). These shift deviations provide the most direct NMR assay for the extent of folding of modified sequences.

The unique structure of the Trp cage motif in EX4 and its apparent lack of interactions with the N terminal half of the

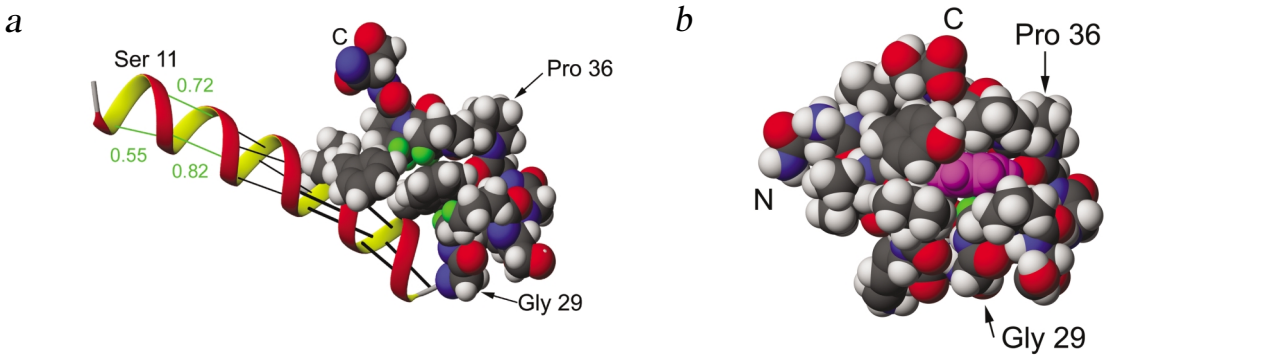

Fig. 1 Truncation and optimization of the C-terminal fold of EX4. ***a***, The structure of EX4 in aqueous 30% TFE. A helix ribbon is shown for residues 8–28, with the side chain atoms of Leu 21, Phe 22, Trp 25 and all atoms of residues 29–39 are displayed with standard CPK color scheme: carbon, black; hydrogen, white; nitrogen, blue; and oxygen, red. The green spheres represent hydrogens that display large upfield shifts due to ring current effects. Hydrogen bonds producing exchange protection are indicated as follows: the thick and thin black lines represent NHs with hydrogen bond fractional populations ≥0.9994 and 0.985–0.998, respectively. Green lines in the substantially frayed N-terminus represent hydrogen bonds with the fractional population shown. ***b***, CPK model of the designed miniprotein TC5b . The indole ring is highlighted in magenta; the strongly upfield-shifted Gly30-Hα2 is shown in green.

Table 1 NMR folding measures for truncated sequences

	Sequence[1]			Folding measures		
	11	21	31	% C-cap[2]	%-cage[3]	Δδ(30-α2)
EX4	---SKQMEEEAVR	LFIEWLKNGG	PSSGAPPPS-NH2	99 [*100*][4]	40 [*84*]	–1.74 [*–3.00*]
TC1a	Ac-**SED**EAVR	LFIEWLKNGG	PSSGAPPPS-NH2	[*95*]	[*85*]	[*–3.03*]
TC1c	**SED**EAVR	LFIEWLKNGG	PSSGA**D**PPS	77 [*95*]	23 [*54*]	–1.37 [*–2.45*]
TC2	**NEVE**	LFI**R**WLKNGG	PSSGAPPPS	[*92*]	[*85*]	[*–3.04*]
TC3b	**N**	LFIEWLKNGG	PSSGAPPPS	61 [*93*]	38 [*86*]	–1.52 [*–2.95*]
TC4a	**D**	LFIEWLKNGG	PSSG**R**PPPS	[*96*]	[*90*]	[*–2.90*]
TC4c	**KG**	LFIEWLKNGG	PSSG**R**PPPS	46 [*92*]	30 [*91*]	–1.09 [*–2.93*]
TC5a	**N**	LFI**Q**WLK**D**GG	PSSG**R**PPPS	63 [*85*]	94 [*94*]	–3.25 [*–3.19*]
TC5b	**N**	L**Y**I**Q**WLK**D**GG	PSSG**R**PPPS	66 [*77*]	100 [*96*]	–3.43 [*–3.28*]

[1]Positions that are mutated *versus* the exendin-4 sequence are shown in bold.
[2]The C-cap measure is the sum of the upfield shifts of 23-α/26-α/30-α3/31-δ3.
[3]%-cage is from the sum of the upfield shifts for the 37-α,-β3/38-δ2-δ3 resonances.
[4]Italic values in brackets are measured in 30% TFE. The 'aqueous' values for EX4 are for 30:70 glycol:buffer (pH 5.9) at 320 K. All other shift measures were obtained at 280 K.

sequence suggest that the domain could be extracted and still form a stable structure. Tests of a series of truncated/mutated sequences (Table 1) in aqueous fluoroalcohol media indicate that all of the species, with the exception of Trp cage construct 1c (TC1c), are ≥97% folded. Folding in water was monitored in several ways: the absence of multiple Xaa Pro *cis*/*trans* isomers, the presence of characteristic long range NOEs and diagnostic chemical shift deviations. Of these, the CSDs provide the more quantitative measures of folding (Table 1). Helicity could be tracked by the Hα CSDs for the helix (residues 21 27), with the shifts of 30 α3 and 31 δ3 monitoring C capping. The sum of the upfield shifts for C terminal Pro resonances (37 αβ3, and 38 δ2 and δ3) monitors Trp cage formation. The largest values observed for the 'C cap' and 'cage formation' measures were equated with '100% folded'. The upfield shift of Gly 30 α2 reflects both C capping and tertiary structure formation.

All of the helix that is not buried by the Pro 37 Pro 38 unit can be eliminated without disrupting the Trp cage, so long as the remaining helix is stabilized by an N capping box (TC1 and TC2) or a single N capping residue (Asp, Asn or Gly). TC1c was the first species prepared that is readily soluble in water but only ~23% folded based on the CSDs for Pro 37 and Pro 38. In the partially folded 20 residue constructs, the extent of tertiary structuring in water correlates with the N cap propensities (Asp ≥ Asn > Gly)[16], and all measures of tertiary structuring and helicity are highly correlated. However, the C terminal Trp cage fold of these 20mers is, at most, 40% populated according to the diagnostic chemical shift criteria.

An examination of the NMR structure (data not shown) derived for TC4a mutant revealed that the Arg 35 side chain was located close to the Asn 28 side chain function. An N28D mutation was incorporated (TC5a and TC5b) in hopes of creating an hydrogen bonded salt bridge. This mutation was coupled with an E24Q mutation to avoid a potentially unfavorable coulombic interaction (EXXXD) in the helix (introducing in its place a pH dependent helix favoring QXXXD interaction)[17]. The F22Y mutation was included in TC5b because Tyr side chains are more commonly observed to have hydrophobic stacking interactions with both Pro and Trp residues in proteins. TC5b is >95% folded in water, displaying the same CD spectrum with or without the addition of TFE (Fig. 2*a*), and melts cooperatively (Fig. 2*b*). The ring current shifts of TC5b are remarkable: Pro 37 β3 (δ = 0.34 p.p.m.) and Gly 30 α2 (δ = 0.72 p.p.m.) are the two furthest upfield resonances in the ^{1}H NMR spectrum (D_2O buffer at pH* 6 (uncorrected for isotope effect)). The shift of Gly 30 α2 seems to be the most upfield observation for a non heme protein. NMR lineshapes and the concentration independence of all spectroscopic parameters imply a monomeric structured state. With TC5b, we appeared to have achieved our stated goal, and further studies focused on this construct.

TC5b fold characteristics and stability

A structure ensemble for TC5b was generated from 169 NOE distances, of which 28 were i/i+n (n > 4) constraints. The key long range NOEs involving the central Trp residue are illustrated in Fig. 3. Even though 10 NHs are exchange protected (*vide infra*), no hydrogen bond constraints are employed because there is no independent basis for assigning the electron pair donors involved. This highly conservative data treatment produced a well converged structure (Fig. 3*a*; Table 2), with an α helix from Leu 21 to Lys 27 and a short 3_{10} helix (residues 30 33). The unusual features are a Gly30 NH hydrogen bond to the i 5 back bone carbonyl, an indole NHε1 hydrogen bond to the i+10 back bone carbonyl and the placement of Pro rings on both faces of the indole ring, with the Tyr ring completing the Trp cage hydrophobic cluster. The degree of side chain rotamer restriction observed in the NMR ensemble argues against a fluxional structure. The TC5b structure is compact and globular (Fig. 1*b*).

For measures of fold stability, we used NH exchange protection. Helix/coil transition algorithms (for example, AGADIR)[18], predict protection factors ≤1.5 for the nine residue (including the capping residues) helix of TC5b. In agreement with this prediction, truncated peptides lacking the C terminal $(Pro)_3$ unit do not display measurable NH protection (data not shown). In contrast, NHε1 of Trp 25 and nine backbone NHs (residues 23 28, 30, 33 and 35) of TC5b are significantly protected (protection factor (PF) = $k_{rc} / k_{obs} = 10^{1.01 \pm 0.23}$ at pH* 3.63). Upon titration to pH* 6.03, the Trp 25 Hε1 protection factor increases from $10^{1.10}$ to $10^{1.83}$. This is attributed to the ionization of Asp 28 and fold stabilization through formation of a salt bridge with Arg 35. The formation of tertiary structures clearly results in enhanced exchange protection. To our knowledge, TC5b is the first monomeric peptide of this length to display measurable exchange protection in water. Upon repeating the pH* 6.03 experiment with the addition of 30% TFE, protection factors increase for example, Trp 25 Hε1 ($t_{1/2}$ = 8 h and PF = $10^{2.84}$) and Leu 26 HN ($t_{1/2}$ = 30 h and PF = $10^{3.15}$) in agreement with the increased T_m observed by CD and NMR (Fig. 2*d*).

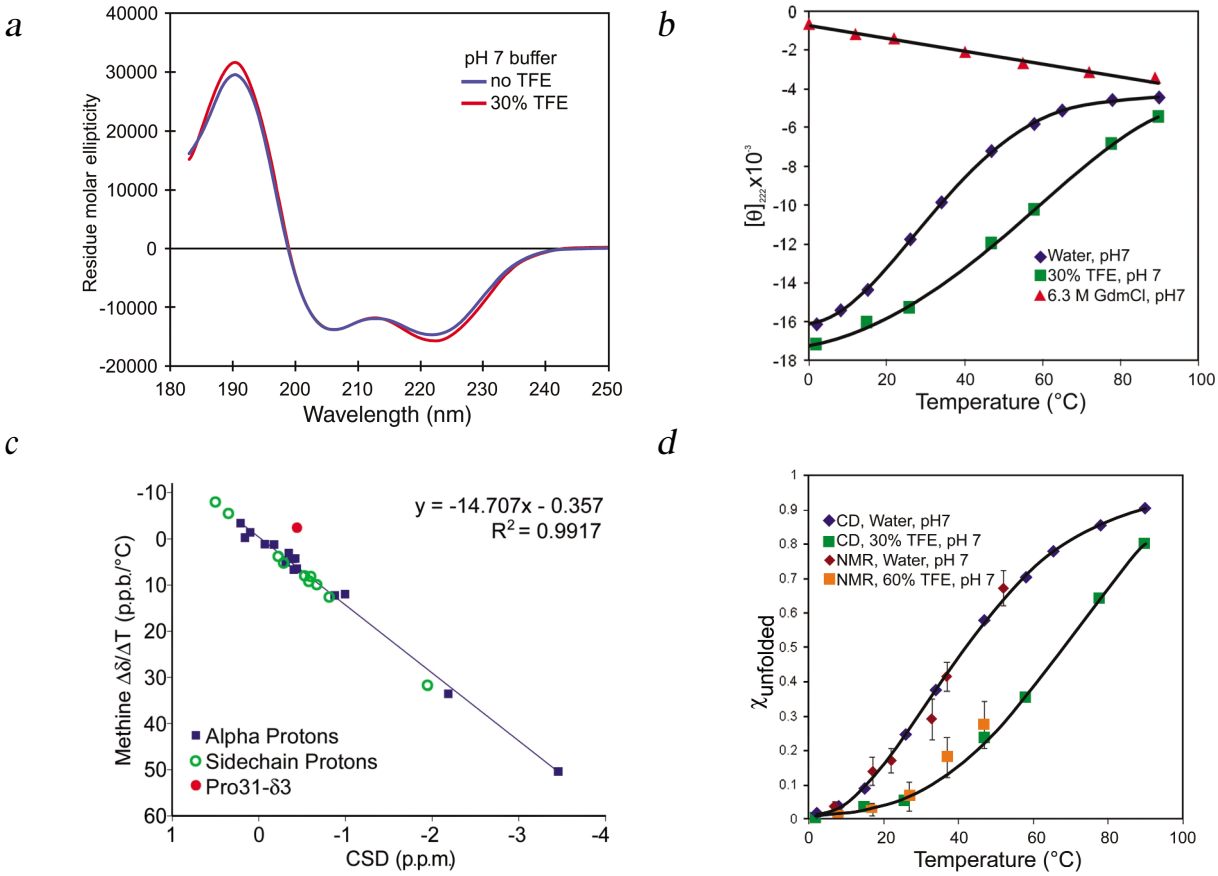

Fig. 2 Folding measures for Trp-cage construct 5b. ***a***, CD spectra of 66 μM TC5b in pH 7 aqueous buffer (2 °C) and buffer with 30% (v/v) TFE: residue molar ellipticity *versus* wavelength. The curve shape suggests an unrealistically high level of helicity from a reinforcing Trp side chain chromophore contribution near the amide n π^* transition. ***b***, CD-monitored melts for TC5b in pH 7 aqueous buffer with and without the addition of 30% TFE or guanidine-HCl (6.3 M): $[\theta]_{222}$ *versus* T (°C). The denatured state data was fit to a line; the curves through the other data points are polynomial fits with no theoretical significance. ***c***, Graphs illustrating the correlation of chemical shift temperature gradients ($\Delta\delta/\Delta T$) with the chemical shift deviations observed in D_2O at 7 °C. All CαH resonances (|CSD| > 0.1 p.p.m.) are shown (solid squares) and are the basis for the least-squares fit; other CH signals (open circles) in the C-terminal portion of the structure fall on the same line. The Pro31-δ3 resonance (red circle) is an outlier (see text). ***d***, Unfolded mole fraction *versus* T (°C) from the CD data and NMR, the latter based on 11 fractional chemical shift deviations (23-α, 26-α, 30-α2; 31-α and -β3; 37-α, -β3 and -δ3; and 38-α, -δ2 and -δ3). The NMR measures are shown with error bars (±s.e.); greater scatter would be expected for non-cooperative unfolding. The CD values were converted to fraction unfolded assuming the guanidine-HCl line in (*b*) represents 100% unfolded and a common temperature gradient ($\partial[\theta]_{222}/\partial T$ = +52° per °C) for the 100% folded baselines. The intercept values of $[\theta]_{222}$ were –16,100° (buffer) and –17,200° (30% TFE).

Trp cage folding cooperativity

We have reported[19] a novel graphical test for folding cooperativity based on CSD values and chemical shift temperature gradients. For two state structure/disorder equilibria, a linear correlation is expected between the CSDs observed at low temperatures and the $\Delta\delta/\Delta T$ values observed during the melting transition; the correlation coefficient is a measure of cooperativity. Excellent correlations are observed for NH shifts (data not shown) and for all Hα resonances, with the ring current shifted resonances appearing on the same line (Fig. 2*c*).

CD studies of melting (Fig. 2*b*) also indicate cooperativity. The melting transition midpoint is 42 °C in pH 7 aqueous buffer; CD measures of helix melting are in complete agreement with NMR CSD melting data from nonhelical portions of the Trp cage (Fig. 2*d*). TC5b also displays the classic hallmarks of cooperative unfolding in a guanidine HCl denaturation titration ($[\text{guanidinium}]_{1/2}$ = 2.2 M at 3 °C and ΔG_U = +8.6 (±0.9) kJ mol ^{1}; data not shown). This corresponds to 97.5% folded, in agreement with the value obtained from the Trp25 NHε1 exchange protection factor (98.5%). The NH protection data, guanidine HCl titration, thermal CD and NMR melting data are in remarkably good agreement, which would be expected only for a two state folding process.

Structure in the absence of the Trp cage

A detailed analysis of thermal chemical shift changes in some of our constructs has revealed two instances of residual structure in states that lack the complete Trp cage. In the earlier constructs with a longer helix (TC1a/c and TC2), the retention of substantial helicity induced shift deviations for all of the Hα resonances in the α helix upon heating is observed in 30% TFE and also (to a lesser extent) in aqueous buffer. The phenomenon is particularly pronounced for EX4. A direct comparison of the exchange rates for EX4 and TC5b in pH* 6.03 buffer with 30% TFE quantified this effect. The Trp 25 NHε1 exchange half lives are 3.0 and 8.0 h for EX4 and TC5b, respectively. For TC5b, only the backbone NH of Leu 26 exchange slower than 25 ε1. In the case of EX4, eight backbone NHs (all in the C terminal section of the α helix) remain after 25 ε1 was >85% exchanged. The backbone exchange rates for the Ile 23 Lys 27

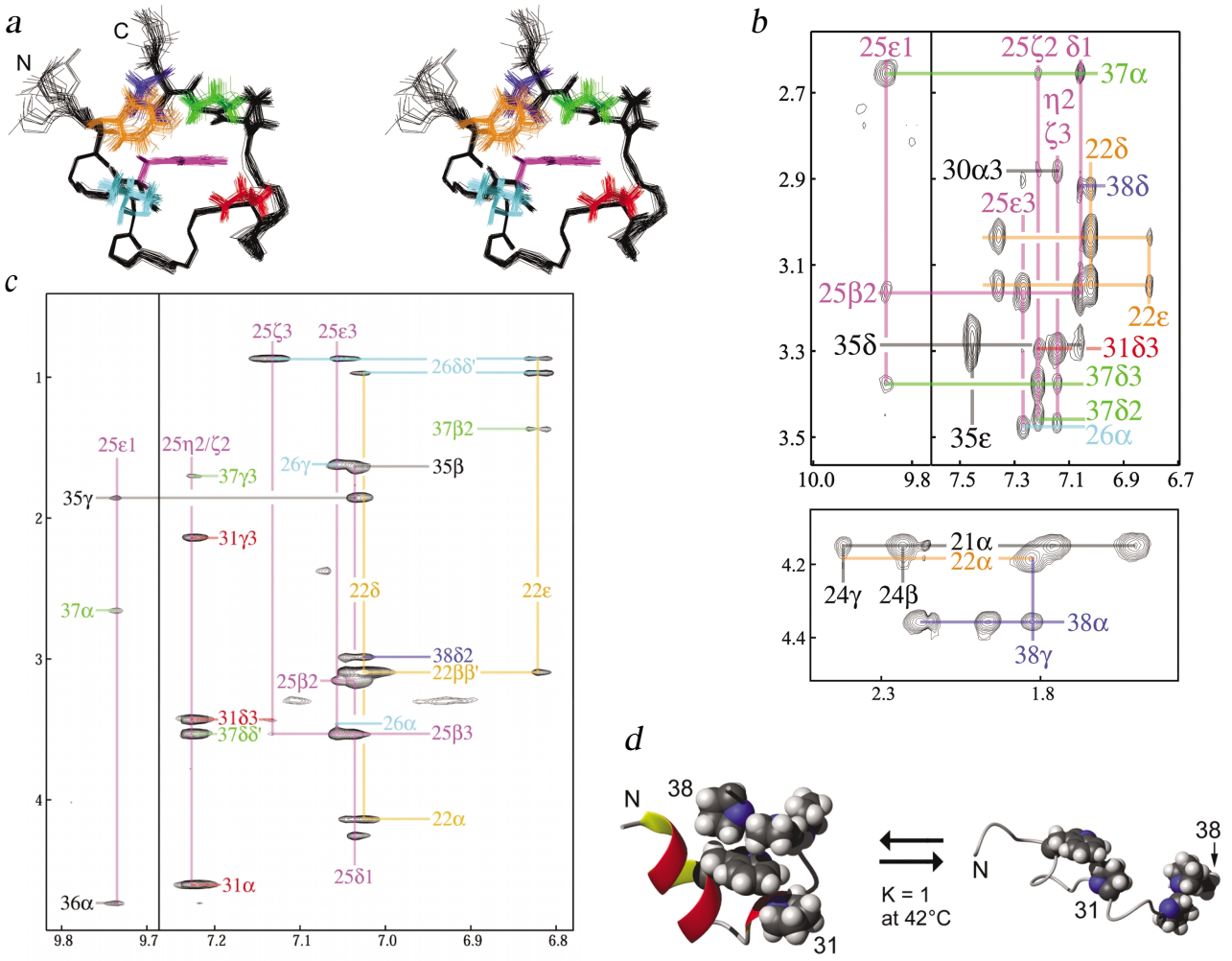

Fig. 3 NMR spectra and the structure derived for TC5b. ***a***, Stereo view of the NMR ensemble (38 of 50 structures for TC5b in pH 7 aqueous buffer (Table 1)). All atoms are displayed for Tyr 22 (orange), Trp 25 (magenta), Leu 26 (cyan), Pro 31 (dark red), Pro 36 (black), Pro 37 (green) and Pro 38 (blue). For the remaining residues, only the backbone is displayed. The heavy-atom pairwise r.m.s. deviation over the key residues in the Trp cage (Tyr 22, Trp 25, Gly 30, Pro 31, Pro 37 and Pro 38) is 0.46 ± 0.15 Å. The annotated NOESY segments ***b***, with added TFE and ***c***, without TFE illustrate the diagnostic long-range NOEs. This is the same color scheme used in (*a*), with Leu 21, Gln 24, Gly 30 and Arg 35 also shown in black. In (*b*), the unlabeled line at 7.36 p.p.m. is Ile 23-HN. The key long-range NOEs of TC5b (for example, 22-α to 38-γ, 22-δ to 38-δ2, 22-ε to 37-β2, 25-ε1 to 35-γ/36-α/37-α, 25-δ1 to 35-β/38-δ2, 25-η2 to 31-δ3 and 25-ζ2 to 31-α/37-δ2-δ3) were observed in both media (the 22-α/38-γ NOE does not appear in (*c*)). Trp 25-Hδ1/Hε3 and Hη2/Hζ2 are nearly and completely shift coincident, respectively, in aqueous buffer. Long-range NOEs to these indole ring resonances are attributed to individual hydrogen sites based on their occurrence in other Trp-cage constructs under conditions where the resonances are not shift coincident. ***d***, TC5b is in a temperature dependent equilibrium with an 'unfolded state' that does not display random coil shifts because of residual local hydrophobic cluster formation. The increasingly negative CSDs of 31-δ3 and 30-α3 are rationalized because the 'residual' high temperature hydrophobic cluster between Trp 25 and Pro 31 places these two protons further into the shielding region than their location in the unmelted Trp cage.

segment of EX4 indicate a PF of $10^{5.1 \pm 0.7}$, which is 500 fold greater than the protection factor of the 25 ε1. Thus, in the case of EX4, Trp cage formation corresponds to the docking of the C terminal $(Pro)_3$ unit onto an exposed Trp side chain of a preformed helix.

In truncated, optimized Trp cage constructs, there is no evidence for retained helicity after thermal loss of tertiary structure; however, there are still some 'unexpected' CSD changes at Gly 30 and Pro 31 resonances upon melting and/or mutation. In the optimized constructs (TC5a/b), the Gly 30 CH_2 displays dramatic stereo selective shielding (3.43 for Hα2 and 0.96 p.p.m. for Hα3), and Pro 31 δ3 is modestly shifted (CSD = 0.26 to 0.47 p.p.m.). In the unoptimized short constructs (and in constructs with a longer α helical segment), 30 α3 CSDs are as large as 1.56 and 31 δ3 CSD values as large as 0.87 p.p.m. were observed. Furthermore, although all of the other Pro δ resonances move uniformly toward their random coil values on melting, warming aqueous solutions of TC5a/b shifts the 31 δ3 resonance farther upfield (the outlier in Fig. 2*c*). A structural model of the unfolding of TC5b rationalizes these observations (Fig. 3*d*). A similar hydrophobic cluster between Trp 25 and Pro 31 may serve to C cap the helix of EX4 in the DPC micelle associated state that lacks tertiary structure[15].

Implications

If the NMR structure ensemble correctly reflects the structure and the motions within the folded state energy, it should predict the chemical shifts. Ring current shift calculations were performed using SHIFTS 3.1 (*http://www.scripps.edu/case/*) and MOLMOL[20]. The structure predicts all of the >0.2 p.p.m.

Table 2 NMR structure statistics[1]

Distance restraints and r.m.s deviations in the CNS ensemble[2]		
Type of restraint	Number	R.m.s. deviation[3]
Intraresidue	43	0.009 ± 0.002
Sequential	62	0.015 ± 0.002
i / i + n, n = 2–4	36	0.009 ± 0.003
i / i + n, n ≥ 5	28	0.010 ± 0.003
Structure statistics[3]		
E_{LJ} (kcal mol^{-1})	–74.7 ± 3.4	
Bond violations (Å)	0.0032 ± 0.0001	
Angle violations (°)	0.45 ± 0.01	
Improper torsion violations (°)	0.16 ± 0.01	
Convergence within final ensemble, atomic r.m.s. deviations (Å)[4]		
Pairwise over the ensemble (±s.e.)		
Backbone	0.41 ± 0.12	
Heavy atom	0.78 ± 0.15	

[1]All statistics are over the best-fit 38 of 50 structures. One of the 50 structure generating runs was a total failure (numerous violations >0.4 Å and E_{LJ} > +70 kcal mol^{-1}). The remaining 11 structures that were excluded had one or more violations >0.16 Å and E_{impr} values that were significantly higher than the remaining 38 structures.
[2]In accepted structures, there were no restraint violations >0.11 Å. The acceptance criteria were no NOE violations >0.15Å and agreement with structural norms. The latter were judged by bond, angle and improper torsion violations, the mean of which could not exceed 0.01 Å, 0.5° and 0.2° in any structure with no individual values exceeding 0.02 Å, 3° and 2°, respectively.
[3]Values are the mean ± standard deviation.
[4]Each accepted structure was briefly minimized (500 steps and steepest descent) in the AMBER 6 (*http://www.amber.ucsf.edu/amber/*) force field. This minimization produced no significant changes in the convergence within or restraint violations of the ensemble. All convergence measures are over residues 21 38, excluding the N- and C-terminal residues. In addition, the heavy atom measure excludes the long side chains of Leu 21, Lys 27 and Arg 35.

CSDs. For large CSDs due primarily to the indole ring, the observed values are within 14 ± 8 % (n = 10) of the SHIFTS 3.1 ring current predictions. Notably the highly stereo specific CSDs of the Gly 30 CαH$_2$ (3.43 and 0.96 p.p.m. observed *versus* 3.17 ± 0.39 and 0.91 ± 0.17 calculated from the ensemble) and the large upfield shifts for Pro 37 Pro 38 sites are reproduced. The agreement obtained using the NMR ensemble is consistent with a structure that has relatively little additional motion and remains tightly packed about the indole ring. We have significantly lowered the size limit for a fully protein like fold.

The potential for a coulombic interaction between Asp 28 and Arg 35 in the folded state is the basis for the N28D mutation. At pH 7, TC5a is 7.6 kJ mol^{-1} more stable than the TC3b. NMR monitored pH titrations of TC5a/b suggest an Asp deprotonation $\Delta\Delta G$ of 5.8 7.5 kJ mol^{-1}. Similar fold stability increments have been observed upon the electrostatic optimization of protein surfaces[21]. We view the sequence remote coulombic interaction as the best rationale; however, additional studies are required to ascertain what portion, if any, of the stabilization is due to the pH dependent, helix favoring QXXXD interaction[17].

The spectroscopic probe (primarily chemical shifts) and co solvents used to monitor stability in this fold minimization and optimization study deserve some comment. CSD analysis seems to be optimal for studying fast folding systems at the peptide/protein borderline. There has been considerable controversy concerning the extent to which the peptide structuring effects of TFE and hexafluoroisopropanol (HFIP)[22,23] are pertinent to the 'native' states of peptides and proteins in water. Fluoroalcohol stabilization of helices has been known for decades, and more recent studies have extended this to β hairpins[14,24,25] and β sheets[10]. With this report, fluoroalcohol induced increases in 'native aqueous structure' populations have been observed for a system that owes its stability to a specific hydrophobic core (the Trp cage).

As previously noted[15], the Trp cage fold is an intramolecular example of a motif for binding Pro rich segments to domains that are involved in signal transduction. Pro Trp interactions may also be an effective strategy for fold stabilization. There is some analogy between the environment of our buried indole ring and that of Trp 11 in the Pin WW domain; Tyr 22 Trp 25 Pro 36 Pro 37 Pro 38 in the Trp cage equates with Tyr 23 Trp 11 Pro 8 Leu 7 Pro 37 in the WW domain[9]. Pro residues are advantageous because their inclusion serves to reduce the entropic advantage of the unfolded state. It is not coincidence that the smallest natural proteins also have Pro residues suspended over Trp side chains. Finally, Trp cage constructs may prove to be a useful paradigm for protein folding studies in which both experiments and computational simulations are simplified by the small size of the structures. As the smallest protein like system known, these systems should be an excellent testing ground for molecular dynamics simulations of protein unfolding and folding pathways.

Methods

Peptide synthesis. The sample of EX4 has been described[15]. All other peptides were prepared using fast FMOC chemistry on an ABI 433A peptide synthesizer and purified by reversed-phase HPLC (C_{18} column) with a water + 0.1% TFA:acetonitrile + 0.085% TFA gradient. Purity and sequence were established from the NMR spectra.

CD spectroscopy and melting studies. CD samples were prepared by dissolving weighed amounts (0.5–2 mg) of lyophilized peptides directly in 15 mM aqueous phosphate (pH 5.9–7.05) or phosphate-acetate (pH 3–4.5) buffer to produce ~600μM stock solutions (determined by UV at $\varepsilon_{278\,nm}$ = 5,580 cm^2 mmol^{-1} for Trp or 6,760 cm^2 mmol^{-1} for Trp + Tyr). Quantitative serial dilutions to the required levels of fluoroalcohol and aqueous buffer were used. CD spectra were recorded in 1 and 10 mm path length cells (25–70 and 2–10 μM peptide concentrations, respectively) using a JASCO model J720 spectropolarimeter as reported[26]. CD spectral values for peptides are expressed in units of residue molar ellipticity (deg cm^2 dmol^{-1}).

NMR spectroscopy and structure ensemble generation. Solution-state NMR samples were made by dissolving 2–3.5 mg lyophilized peptide in 450–500 μl aqueous buffer (as listed in CD methods) with 10% D_2O for locking. Structuring shifts are reported as CSDs, experimental shift/random coil values[19], with referencing to internal DSS. NOESY and TOCSY spectra were collected for all media, with solvent suppression accomplished using the WATERGATE pulse sequence[27]. All the spectra were collected on a Bruker DRX operating at 500 MHz and were readily and unambiguously assigned by standard methods[28]. The complete chemical shift assignments for TC5b in aqueous buffer with and without added TFE have been deposited (BMRB entry 5292). NOE intensities were converted to Å distances — short (2.0–3.0 Å), medium (2.5–3.5 Å), long (2.9–4.0 Å) or very long (3.3–5.0 Å) — and used in CNS[29] protocol as described[15]. The weighting was 75 kcal Å^{-2} during the final minimization, which included a Lennard-Jones rather than purely repulsive van der Waals terms. The acceptance criteria, violations (none) and convergence statistics appear in Table 2. All structural figures in this report were prepared using MOLMOL[20].

NH exchange studies and protection factor (PF) analysis. Exchange rates were obtained from 1D spectra as the slopes of plots of ln (NH signal intensity) *versus* time. The experiments were performed at 7–9 °C by adding pre-cooled D_2O buffer to the lyophilized peptide sample in a pre-cooled NMR tube. After brief vortex mixing, the sample was placed in the previously cooled and shimmed NMR probe for immediate data collection. Additional points were recorded first at 5–15 min intervals and then daily for long exchange times. Assignments for NHs that were still present 2–4 h after dissolution in D_2O buffer were confirmed by 2D correlations. PF values were obtained using the protocols and coil reference values used for EX4 (ref. 15).

Coordinates. Coordinates and NMR distance constraints have been deposited in the Protein Data Bank (accession code 1L2Y).

Acknowledgments
Initial support came from a feasibility grant from the University of Washington Royalty Research Fund with continuing support from an NIH grant. We thank L. Serrano (EMBL-Heidelberg) for reminding us of the pH dependence of the helix-favoring QXXXD interaction.

Competing interests statement
The authors declare that they have no competing financial interests.

Correspondence should be addressed to N.H.A. *email: andersen@chem.washington.edu.*

Received 19 November, 2001; accepted 28 March, 2002.

1. Dahiyat, B.I. & Mayo, S.L. *Science* **278**, 82 87 (1997).
2. Hill, R.B. & DeGrado, W.F. *J. Am. Chem. Soc.* **120**, 1138 1145 (1998).
3. Walsh, S.T.R., Cheng, H., Bryson, J.W., Roder, H. & DeGrado, W.F. *Proc. Natl. Acad. Sci. USA* **96**, 5486 5491 (1999).
4. Ottesen, J.J. & Imperiali, B. *Nature Struct. Biol.* **8**, 535 539 (2001).
5. Cochran, A.G., Skelton, N.J. & Starovasnik, M.A. *Proc. Natl. Acad. Sci. USA* **98**, 5578 5583 (2001).
6. Li, X., Sutcliffe, M.J., Schwartz, T.W. & Dobson, C.M. *Biochemistry* **31**, 1245 1253 (1992).
7. Sudol, M. *Prog. Biophys. Mol. Biol.* **65**, 113 132 (1996).
8. McKnight, C.J., Doering, D.S., Matsudaira, P.T. & Kim, P.S. *J. Mol. Biol.* **260**, 126 134 (1996).
9. Jager, M., Nguyen, H., Crane, J.C., Kelly, J.W. & Gruebele, M. *J. Mol. Biol.* **311**, 373 393 (2001).
10. Kortemme, T., Ramírez Alvarado, M. & Serrano, L. *Science* **281**, 253 256 (1998).
11. López de la Paz, M., Lacroix, E., Ramírez Alvarado, M. & Serrano, L. *J. Mol. Biol.* **312**, 229 246 (2001).
12. Schenck, H. & Gellman, S. *J. Am. Chem. Soc.* **120**, 4869 4870 (1998).
13. Maynard, A.J., Sharman, G.J. & Searle, M.S. *J. Am. Chem. Soc.* **120**, 1996 2007 (1998).
14. Andersen, N.H. *et al. J. Am. Chem. Soc.* **121**, 9879 9880 (1999).
15. Neidigh, J.W., Fesinmeyer, R.M., Prickett, K.S. & Andersen, N.H. *Biochemistry* **40**, 13188 13200 (2001).
16. Andersen, N.H. & Tong, H. *Protein Sci.* **6**, 1920 1936 (1997).
17. Huyghues Despointes, B.M., Klinger, T.M. & Baldwin, R.L. *Biochemistry* **34**, 13267 13271 (1995).
18. Muñoz, V. & Serrano, L. *Biopolymers* **41**, 495 509 (1997).
19. Andersen, N.H. *et al. J. Am. Chem. Soc.* **119**, 8547 8561 (1997).
20. Koradi, R., Billeter, M. & Wüthrich, K. *J. Mol. Graph.* **14**, 51 55 (1996).
21. Loladze, V.V., Ibarra Molero, B., Sanxhez Ruiz, J.M. & Makhatadze, G.I. *Biochemistry* **38**, 16419 16423 (1999).
22. Andersen, N.H., Cort, J.R., Liu, Z., Sjoberg, S.J. & Tong, H. *J. Am. Chem. Soc.* **118**, 10309 10310 (1996).
23. Walgers, R., Lee, T.C. & Cammers Goodwin, A. *J. Am. Chem. Soc.* **120**, 5073 5079 (1998).
24. Blanco, F.J. & Serrano, L. *Eur. J. Biochem.* **230**, 634 649 (1995).
25. Ramírez Alvarado, M., Blanco, F.J. & Serrano, L. *Protein Sci.* **10**, 1381 1392 (2001).
26. Andersen, N.H., Liu, Z. & Prickett, K.S. *FEBS Lett.* **399**, 47 52 (1996).
27. Piotto, M., Saudek, V. & Sklenar, V. *J. Biomol. NMR* **2**, 661 665 (1992).
28. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (John Wiley, New York; 1986).
29. Brünger, A.T. *et al. Acta Crystallogr. D* **54**, 905 921 (1998).