Protein Science (1997), *6*:1869–1877. Cambridge University Press. Printed in the USA.

Solution structure of $\alpha t\alpha$, a helical hairpin peptide of de novo design

YOUCEF FEZOUI,[1] PETER J. CONNOLLY, AND JOHN J. OSTERHOUT

The Rowland Institute for Science, Cambridge, Massachusetts 02142

(RECEIVED March 5, 1997; ACCEPTED May 12, 1997)

Abstract

$\alpha t\alpha$ is a 38-residue peptide designed to adopt a helical hairpin conformation in solution (Fezoui Y, Weaver DL, Osterhout JJ, 1995, *Protein Sci 4*:286–295). A previous study of the carboxylate form of $\alpha t\alpha$ by CD and two-dimensional NMR indicated that the peptide was highly helical and that the helices associated in approximately the intended orientation (Fezoui Y, Weaver DL, Osterhout JJ, 1994, *Proc Natl Acad Sci USA 91*:3675–3679). Here, the solution structure of $\alpha t\alpha$ as determined by two-dimensional NMR is reported. A total of 266 experimentally derived distance restraints and 20 dihedral angle restraints derived from J-couplings were used. One-hundred initial structures were generated by distance geometry and refined by dynamical simulated annealing. Twenty-three of the lowest-energy structures consistent with the experimental restraints were analyzed. The results presented here show that $\alpha t\alpha$ is comprised of two associating helices connected by a turn region.

Keywords: α-helix; de novo design; NMR solution structure; peptide

Protein folding is usually envisioned to proceed through a series of steps starting with the formation of individual elements of secondary structure (or the formation of microdomains) and proceeding through stages where these structures come together, interact, and become mutually stabilizing (Karplus & Weaver, 1976, 1994; Kim & Baldwin, 1982; Baldwin, 1989, 1995). Eventually, the final protein structure is attained. Because protein folding intermediates are only transiently populated, the study of their structures and stabilities is problematic.

In order to circumvent these difficulties, peptides have been used as models for the early protein folding intermediates. Much of the work has addressed stability issues in helical peptides (for reviews see Scholtz & Baldwin, 1992; Chakrabartty & Baldwin, 1995), although turn peptides (Dyson et al., 1988a) and template-initiated sheet peptides (Kemp, 1990) have also been studied.

The next logical step is to extend the field of inquiry to linked secondary structures in order to access the effects of mutual stabilization. This area has received much less attention. A model system consisting of disulfide-linked peptides making up the core of bovine pancreatic trypsin inhibitor has been developed (Oas & Kim, 1988) and a related system was used to study the effects of charge (Kwon & Kim, 1994a) and hydrophobicity (Kwon & Kim, 1994b) on peptide stability. The stability of the isolated α-helix from this system has also been examined (Goodman & Kim, 1989). Linked helices with a minimalist sequence have been used to investigate salt- and pH-induced conformational changes (Goto & Aimoto, 1991). Hodges and coworkers have studied minimalist coiled coils extensively (see Adamson et al., 1993), although this system was intended as a simple model of protein structure rather than of protein folding intermediates.

$\alpha t\alpha$ was developed to provide a model system for a protein folding intermediate at the level of secondary structure association (Fezoui et al., 1995a). The sequence of $\alpha t\alpha$ was designed de novo in order to circumvent problems with aggregation of helix-turn-helix peptides derived from natural proteins and to allow the development of a sequence that would be tractable for NMR assignment and structure determination.

Structure verification has been a particular problem for de novo-designed peptides and proteins. The advantages of the α_1B family of proteins, minimalist sequence and symmetry, work against the structure determination of these molecules by NMR (Ciesla et al., 1991; Osterhout et al., 1992). Although the X-ray structure of one of this family of molecules, α_1A, is consistent with certain aspects of the original design, it is not clear that the crystal structure reflects the solution structure of the molecule accurately (Hill et al., 1990).

Another problem with designed molecules that works against structural determination is that they often exhibit certain properties of the molten globule state of proteins (Ohgushi & Wada, 1983;

Reprint requests to: John J. Osterhout, The Rowland Institute for Science, 100 Edwin H. Land Blvd., Cambridge, Massachusetts 02142; e-mail: osterhout@rowland.org.

[1] Present address: Department of Neurology, Harvard Medical School and Center for Neurologic Diseases, Brigham and Women's Hospital, 77 Avenue Louis Pasteur (HIM 750), Boston, Massachusetts 02115.

Abbreviations: NOESY, NOE spectroscopy; TOCSY, total correlation spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; RMSD, RMS deviation; VDW, van der Waals.

Dolgikh et al., 1985). These molecules might be termed more properly as gemish (mixed) states because they fold to an ensemble of conformations rather than molten globule states, which are altered states of globular proteins (Dill et al., 1995). α_1B (Osterhout et al., 1992; Handel et al., 1993) and Felix (Hecht et al., 1990) are de novo-designed molecules that are gemish states. Variants of α_1B have exhibited more native protein-like properties (for instance, sigmoidal melting curves) (Raleigh & DeGrado, 1992; Raleigh et al., 1995) and a de novo-designed β-sandwich protein also exhibits sigmoidal melting (Quinn et al., 1994; Yan & Erickson, 1994). These molecules are not necessarily protein-like, because their melting curves are broader than is typical of globular proteins and the amide protein exchange rates have not been measured.

Successful structure determination for de novo-designed molecules recently has been the provenance of relatively small peptides. NMR solution structures have been obtained for BB, a 12-residue β-hairpin peptide (Sieber & Moe, 1996); ALIN, a 35-residue α-helical hairpin peptide (Kuroda et al., 1994); and BBA1, a 23-residue $\beta\beta\alpha$ peptide (Struthers et al., 1996). In an earlier study of the carboxylate form of $\alpha t\alpha$, NOEs were observed that were consistent with the designed tertiary structure (Fezoui et al., 1994, 1995b). The present paper concerns the determination of tertiary structure of $\alpha t\alpha$ by two-dimensional NMR.

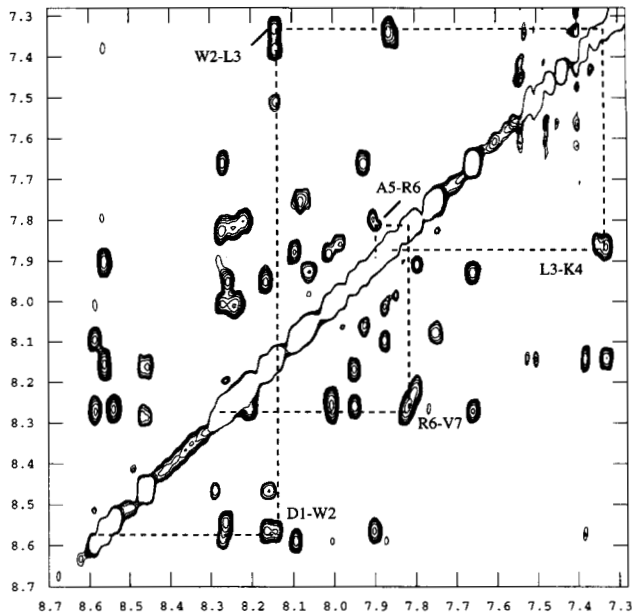

Fig. 1. Contour plot of the amide–amide region of a 200-ms mixing time NOESY of $\alpha t\alpha$. Solution conditions were 3 mM $\alpha t\alpha$ at 25 °C, pH 3.6, and 20 mM NaCl. Select amide–amide cross peaks are indicated. This figure was produced with the Rowland NMR toolkit (Hoch & Stern, 1993; http://www.rowland.org/rnmrtk/toolkit.html).

Results

NMR assignments

The sequential NMR assignment of $\alpha t\alpha$ was accomplished using standard two-dimensional NMR techniques (Wüthrich, 1986). The analysis program EASY (Eccles et al., 1991) was used during the assignment process. Spin systems were assigned using DQF-COSY and TOCSY spectra. Eight of the 38 amino acids of $\alpha t\alpha$ are unique. These are Trp 2, Val 7, Gly 18, Thr 19, Ser 21, Asn 22, Met 28, and Ile 36. The spin systems of these amino acids were identified easily and were used as starting points for the sequential assignment of the remaining residues. Sequential assignments were determined through identification of short-range $d_{\alpha N}(i, i + 1)$, $d_{NN}(i, i + 1)$, and $d_{\beta N}(i, i + 1)$ connectivities in the NOESY spectrum. Examples of sequential $d_{NN}(i, i + 1)$ connectivities are shown in Figure 1. Chemical shift assignments for $\alpha t\alpha$ at pH 3.6 and 25 °C are presented in Table 1.

Identification of secondary structures

The sequential and medium-range NOEs, $^3J_{HN\alpha}$ coupling constants, and the α-proton chemical shift index (Wishart et al., 1992) are summarized in Figure 2. Overlapping medium-range $d_{\alpha N}(i, i + 3)$, $d_{\alpha\beta}(i, i + 3)$, and $d_{\alpha N}(i, i + 4)$ NOEs, which are characteristic of helices, extend from the amino-terminal residue, Asp 1, to Arg 17, the last residue before the designed turn region. Short-range $d_{NN}(i, i + 1)$, also characteristic of helices, occur from Asp 1 to Arg 17, except between Glu 15 and Ala 16, which was not observable due to spectral overlap. Although the short- and medium-range NOEs are consistent with helix formation between Asp 1 and Arg 17, the alpha amide coupling constant of Arg 17 is greater than 6 Hz, which indicates that this residue might be significantly less helical than the preceding residues. Further, the α-proton chemical shift index of both Ala 16 and Arg 17 is more characteristic of random coil values than of α-helix, suggesting that the first helix frays slightly near the turn region.

The second helix is also well defined by short- and medium-range NOEs. Overlapping $d_{\alpha N}(i, i + 3)$, $d_{\alpha\beta}(i, i + 3)$, and $d_{\alpha N}(i, i + 4)$ NOEs are observed from Ser 21 to Lys 38 and short-range $d_{NN}(i, i + 1)$ NOEs are also observed for this region, except for two NOEs, Ser 21–Asn 22 and Leu 32–Lys 33, which are obscured by spectral overlap. Although this NOE pattern indicates that the second helix might begin as early as Ser 21, the coupling constants of both Ser 21 and Asn 22 are in the 6–8 Hz range and the α-proton chemical shift index of these residues is near zero. These data suggest that Ser 21 and Asn 22 are less helical than the remainder of the helix. The intensity of the $d_{NN}(i, i + 1)$ NOEs tapers off near the carboxyl terminus of the molecule, indicating that the helix may fray near the last two or three residues.

Three $d_{\alpha N}(i, i + 2)$ NOEs are observed in the helical regions of $\alpha t\alpha$. The distance between these protons in α-helices is 4.4 Å and in 3_{10}-helices is 3.8 Å (Wüthrich, 1986). NOEs of this type are often observed in protein helices and have also been observed in stable monomeric peptides, which are believed to contain significant populations of 3_{10}-helix (Millhauser et al., 1997). Nascent helices, which are considered to consist of a series of turn-like segments, also have also exhibited $d_{\alpha N}(i, i + 2)$ NOEs (Dyson et al., 1988b). Because $d_{\alpha N}(i, i + 2)$ NOEs are observed in proteins and peptides that represent a wide range of structural organization, it is difficult to attach particular significance to their observation in the helical regions of $\alpha t\alpha$.

The turn region shows NOEs consistent with a mixture of turn types. A $d_{\alpha N}(i, i + 2)$ NOE is observed between residues Thr 19 and Ser 21. This type of NOE is observed between residues 2 and 4 of both type I and type II turns, and its observation is consistent with a propensity toward the formation of a turn made up of residues Gly 18–Ser 21, as was designed originally. However, other NOEs observed in this region are not fully consistent with the presence of either a type I or type II beta turn. Type I turns

Table 1. *Chemical shift assignments for αtα at 25 °C and pH 3.6*

Residue	H^N	H^α	H^β	Others
Succ				CH2: 2.502, 2.440
1 Asp	8.560	4.397	2.718, 2.668	
2 Trp	8.123	4.422	3.404, 3.337	H^δ: 7.365 H^ϵ: 7.495, $H^{\zeta}3$: 7.073 H^η: 7.184, $H^{\zeta}2$: 7.457, $H^\epsilon 1$: 10.316
3 Leu	7.308	3.834	1.523, 1.123	H^γ: 1.267, H^δ: 0.814, 0.782
4 Lys	7.847	3.778	1.804	H^γ: 1.443/1.246, H^δ: 1.610
5 Ala	7.886	4.074	1.468	
6 Arg	7.804	4.067	1.838	H^γ: 1.560, H^δ: 3.083
7 Val	8.259	3.701	2.085	H^γ: 0.980, 0.908
8 Glu	8.218	3.985	2.192	H^γ: 2.308
9 Gln	7.996	4.075	2.204	H^γ: 2.550, 2.417, H^δ: 7.463, 6.841
10 Glu	8.252	4.101	2.113	H^γ: 2.550
11 Leu	8.574	4.026	1.738	H^γ: 1.647 H^δ: 0.833
12 Gln	8.083	3.988	2.151	H^γ: 2.533, 2.420, H^δ: 7.417, 6.791
13 Ala	7.850	4.186	1.498	
14 Leu	7.990	4.066	1.841	H^γ: 1.746, H^δ: 0.855
15 Glu	8.256	4.101	2.110	H^γ: 2.600
16 Ala	7.799	4.258	1.484	
17 Arg	7.735	4.313	1.968, 1.903	H^γ: 1.807, 1.732, H^δ: 3.201
18 Gly	8.067	3.981 4.101		
19 Thr	8.145	4.313	4.283	H^γ: 1.202
20 Asp	8.450	4.686	2.803, 2.759	
21 Ser	8.268	4.417	3.968, 3.885	
22 Asn	8.250	4.581	2.928, 2.877	
23 Ala	8.533	4.030	1.470	
24 Glu	8.242	4.153	2.181	H^γ: 2.458
25 Leu	8.190	4.116	1.773	H^γ: 1.662, H^δ: 0.915, 0.874
26 Arg	8.247	4.032	1.907, 1.850	H^γ: 1.702, 1.630, H^δ: 3.242, 3.181
27 Ala	7.935	4.237	1.537	
28 Met	8.151	4.190	2.245, 2.120	H^γ: 2.665, 2.549
29 Glu	8.551	3.893	2.136	H^γ: 2.531, 2.330
30 Ala	7.891	4.086	1.516	
31 Lys	7.783	4.108	1.948	H^γ: 1.545, H^δ: 1.64 H^ϵ: 2.943
32 Leu	8.200	4.053	1.828	H^γ: 1.798, H^δ: 0.862
33 Lys	8.250	3.876	1.865	H^γ: 1.530, 1.364, H^δ: 1.64, H^ϵ: 2.894
34 Ala	7.643	4.202	1.482	
35 Glu	7.912	4.133	2.254, 2.150	H^γ: 2.601, 2.447
36 Ile	8.045	3.902	1.934	$H^{\gamma 1}$: 1.642, 1.133, $H^{\gamma}2$: 0.90, H^δ: 0.832
37 Gln	7.967	4.171	2.106	H^γ: 2.42, H^δ: 7.443, 6.787
38 Lys	7.843	4.178	1.845, 1.811	H^δ: 1.66, H^ϵ: 2.916
$-NH_2$				7.336, 7.160

should have strong $d_{NN}(i, i + 1)$ and medium $d_{\alpha N}(i, i + 1)$ NOEs between positions 2 and 3. Type II turns should have weak $d_{NN}(i, i + 1)$ and strong $d_{\alpha N}(i, i + 1)$ NOEs between these positions (Wüthrich, 1986). In *αtα*, a weak $d_{NN}(i, i + 1)$ NOE and a medium $d_{\alpha N}(i, i + 1)$ are observed between Thr 19 and Asp 20. These NOE data suggest conformational averaging and a mixture of turn types.

Tertiary structure determination

Helix association was indicated by the observation of a series of interhelical long-range NOEs that served to define interactions between all four turns of helix. The first three sets of interfacing hydrophobic side chains (starting from the ends of helices farthest from the turn) were defined by observation of NOEs directly between their side chains. Association of the first pair of hydrophobic side chains, Leu 3–Ile 36, was indicated by the observation of three NOEs: Leu 3 Hδ–Ile 36 Hα, Leu 3 Hδ–Ile 36 Hγ, and Lys 4 Hδ–Ile 36 Hγ. Interaction of the second pair, Val 7–Leu 32, was defined by two NOEs: Val 7 Hγ–Leu 32 Hα and Val 7 Hγ–Leu 32 Hβ. The Leu 11–Met 28 pair was delineated by an NOE between Leu 11 Hδ and Met 28 Hγ. A direct NOE between the side chains of the last pair, Leu 14–Leu 25, was not observed, possibly due to spectral overlap in this region. Instead, an NOE between Glu 15 Hβ and Leu 25 Hγ was observed.

Tertiary structures were calculated using distance and dihedral angle constraints as described in Materials and methods. Initially,

Fig. 2. Diagram of $^3J_{HN\alpha}$, NOE connectivities, and alpha proton chemical shift indices. Coupling constants are represented by closed bars (<6 Hz), grey bars (>6 Hz) and open bars (overlapped). In the NOE section, * and open bars represent NOEs that could not be observed because of spectral overlap. This figure was prepared with the program Vince by Eric LaRosa and Jeff Hoch [part of the Rowland NMR toolkit (Hoch & Stern, 1993) http://www.rowland.org/rnmrtk/vince.html].

100 structures were calculated. The 23 lowest energy structures that have no distance constraint violations greater than 0.5 Å or dihedral angle violations greater than 5° are shown in Figure 3.

A summary of RMSDs and constraint violations is presented in Table 2. The average deviation for the backbone atoms is 1.08 Å and for heavy atoms, 1.77 Å. When residues 18–21 (the central part of the turn region) are excluded from the overlay, the RMSDs drop to 0.99 Å and 1.69 Å for backbone and heavy atoms, respectively. RMSDs as a function of residue number are shown in Figure 4. The central sections of each helix have the lowest RMSDs, then increase toward the termini, suggesting helix fraying (Fig. 4). The highest RMSDs are in the turn region and could be the result of conformational averaging or the lack of observable NOEs arising from spectral overlap or topology. In Figure 3, the 23 lowest-energy structures of $\alpha t\alpha$ were overlaid. The helical regions of these structures align well and the heterogeneity observed in the turn region and near the ends of the helices can be discerned.

The ensemble of structures shown in Figure 3 demonstrates that the basic elements of the design, associating helices connected by a turn region, have been achieved. The amino-terminal helix includes residues 1–17 and the carboxyl-terminal helix extends from residues 23–38. Both helices show some variability in the hydrogen bonding pattern of the terminal amino acids from structure to structure. For instance, in some structures, Arg 17 forms an $(i, i + 3)$ hydrogen bond with Leu 14, which is characteristic of 3_{10} helix. Although Asn 22 was included in the C-terminal helix in the original design, this residue was not part of the C-terminal helix in the 23 low-energy structures. The turn region encompasses roughly residues 18–22 and shows a variety of conformations.

A key element of the design of $\alpha t\alpha$ was the pattern of side-chain interactions in the hydrophobic interface. A stereo view of the superposed structures of $\alpha t\alpha$ showing the backbone and hydrophobic side-chain atoms is presented in Figure 5. The packing pattern of hydrophobic amino acids can be expressed in terms of the relative orientation of the amino acid side chains. The designed side-chain orientation is FBBF, where Leu 14, Leu 11, Val 7, and Leu 3 are forward, back, back, and forward of Leu 25, Met 28, Leu 32, and Ile 36, respectively (see Fig. 6A). It is clear from

Table 2. *Summary of the statistical analysis of the simulated annealing structures*

Residues	All heavy atoms (Å)	Backbone heavy atoms (Å)
Average RMSDs between the 23 refined structures and their average structure		
1–38	1.77	1.08
1–17, 22–38	1.99	0.99
4–16, 24–34	1.48	0.81
RMSDs		
NOEs (Å)	0.030 ± 0.006	
Bonds (Å)	0.0016 ± 0.0003	
Angles (deg)	0.170 ± 0.03	
Improper (deg)	0.190 ± 0.02	
Number of distance violations > 0.3 Å		2.1 ± 0.7

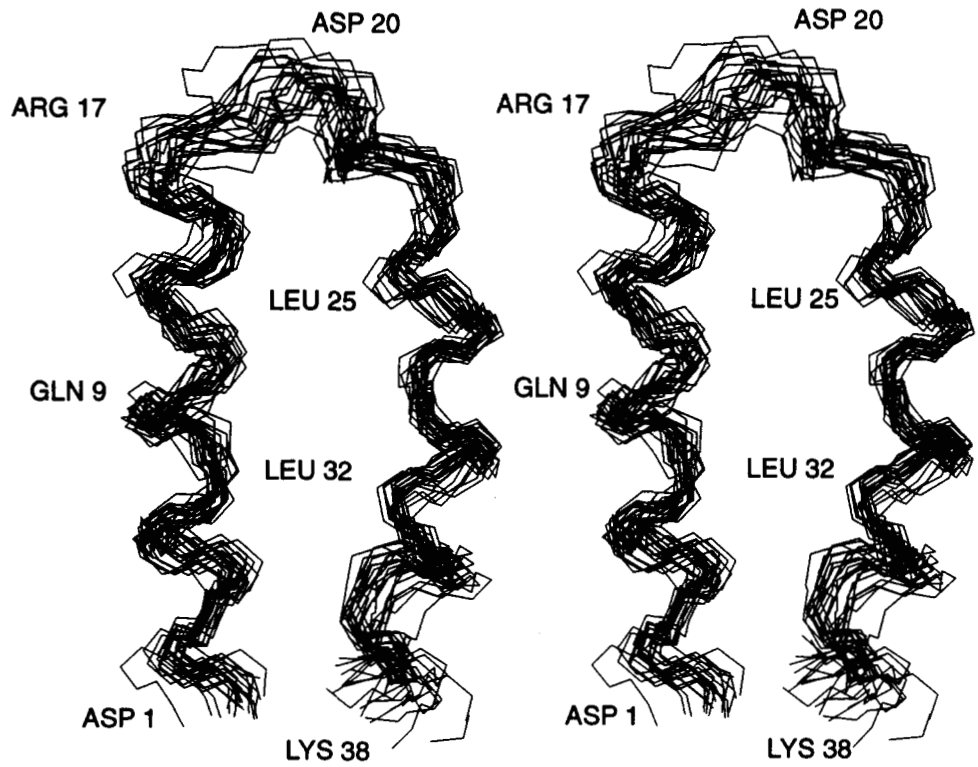

Fig. 3. Stereo view of N, C^{α}, and C′ superposition of the 23 final structures of αtα. Structures were aligned using the alpha carbons of residues 2–17 and 22–37. Prepared in part with MOLMOL (Koradi et al., 1996).

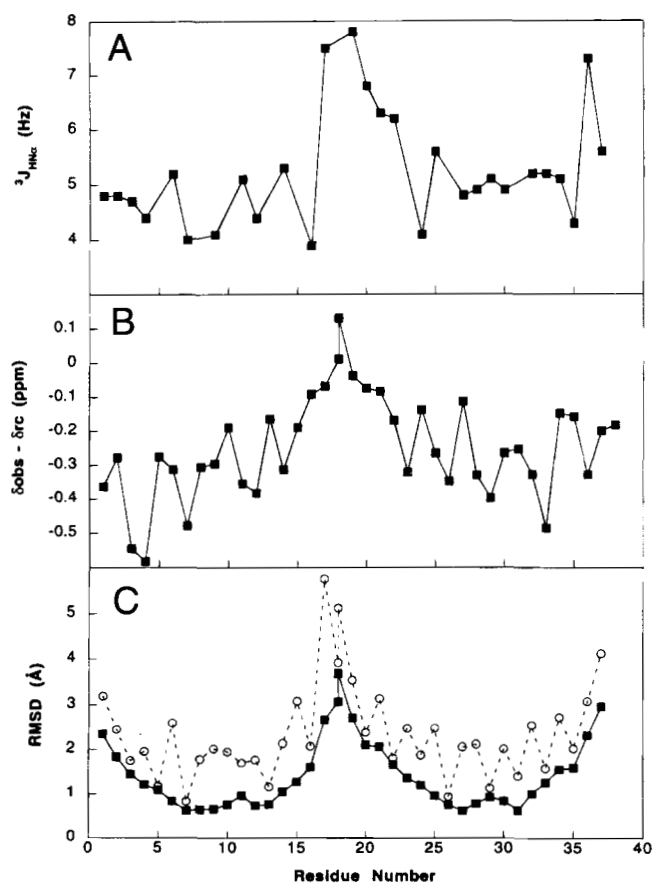

Figure 5 that this pattern is not observed in the ensemble of structures. Most of the observed patterns are of the form FXFF, where X is either B, for back, or E, for even (approximately). The breakdown of patterns is as follows: FBFF (8 structures), FFFF (11 structures), FEFF (1 structure), and FEEF (1 structure—the exception to the FXFF patterns mentioned earlier).

Discussion

The last step of a de novo design cycle for a peptide or protein should be the structural characterization of the molecule by either X-ray diffraction or NMR methods. In earlier reports (Fezoui et al., 1994, 1995b), the observation of certain NOEs consistent with the designed structure of αtα-carboxylate (a variation of αtα that is not amidated at the carboxyl terminus) was reported. In the present study, the three-dimensional solution structure of αtα was determined.

Before examining the correspondence between the designed and solution structures, it is necessary to consider how well the derived

Fig. 4. $^3J_{HN\alpha}$, alpha proton chemical shift differences, and RMSD for αtα. **A:** $^3J_{HN\alpha}$ in Hz plotted against residue number. Average error in these measurements is ±0.85 Hz. Not all coupling constants could be measured due to spectral overlap. Points have been connected to aid in comparison with the other measurements. **B:** Alpha proton chemical shift difference plotted in ppm as the difference between the observed and random coil (Wüthrich, 1986) chemical shifts (δobs − δrc). **C:** RMS difference (Å) relative to the average structure for the final 23 structures of αtα. Closed squares are backbone differences and open circles are to all heavy atoms. Molecules were aligned as described in the legend to Figure 3.

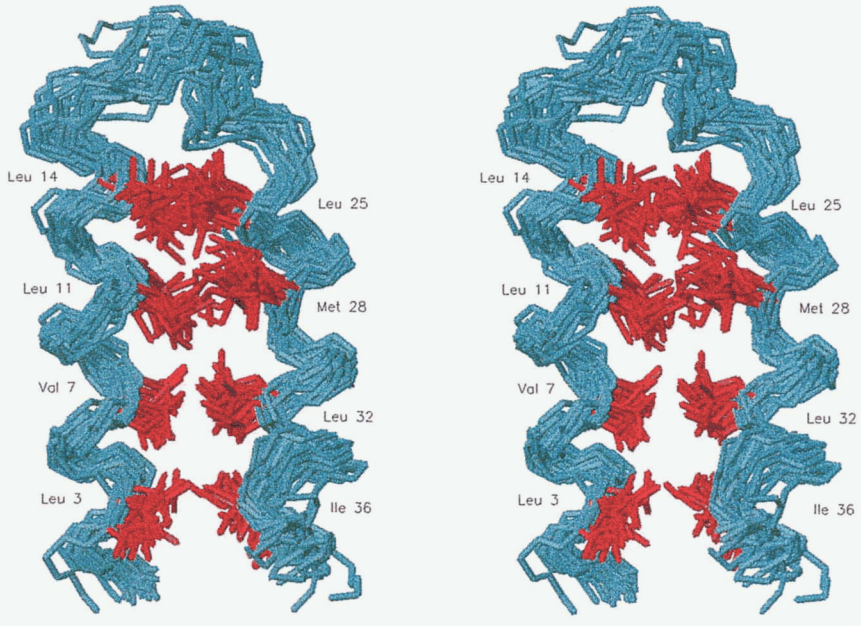

Fig. 5. Stereo view of the superposition of the 23 final structure of $\alpha t\alpha$ showing the backbone atoms (N, C^{α}, and C′) as blue cylinders and the carbon and sulfur atoms of the side chains of the hydrophobic interface as red cylinders. Residue numbers of the hydrophobic interface are shown. Structures were aligned as described in Figure 3. Prepared with MOLMOL (Koradi et al., 1996).

NMR structure reflects the solution structure of the peptide. Short linear peptides in solution are known to be ensembles of rapidly interconverting structures (Wright et al., 1988; Osterhout et al., 1989). It might be expected that $\alpha t\alpha$, as a model for an early folding intermediate, would also exhibit multiple conformations.

Structure determination in the face of conformational averaging is difficult. NOE intensities are $1/r^6$ and population weighted averages of interproton distances. A number of approaches have been developed to handle conformational averaging in NMR structure determination. These approaches employ various methods for cal-

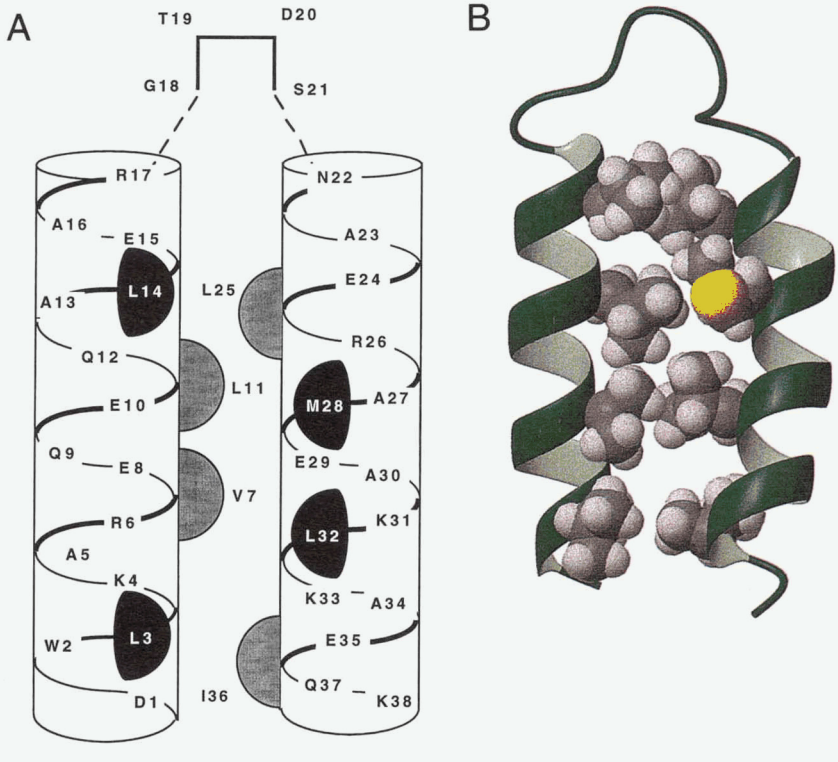

Fig. 6. Designed structure and final structure of $\alpha t\alpha$. **A:** Blueprint of the $\alpha t\alpha$ design adapted from Figure 3 in Fezoui et al. (1995a). **B:** Ribbon/CPK rendering of the refined $\alpha t\alpha$ structure closest to the average structure. Backbone atoms are represented by a ribbon and side chains of the hydrophobic interface are rendered as CPK models. Prepared with MOLMOL (Koradi et al., 1996).

culating a series of structures that are then used to fit the experimental data. Although these approaches can give significantly better fits to the data, a recent study using synthetic data suggests that, even with very accurate NOE distance constraints, it is difficult to determine correctly the relative populations of conformations in multiconformer structures (Bonvin & Brünger, 1996). Here, structures are determined based on a single conformer model, then analyzed in terms of potential conformational averaging considering associated data such as coupling constants and alpha proton chemical shifts.

Short linear peptides are usually ensembles of conformations in solution (Wright et al., 1988). Conformational averaging can arise for a variety of reasons including: folded/unfolded equilibrium, salt bridges (Osterhout et al., 1989), hydrophobic clusters (Bundi et al., 1976, 1978; Wright et al., 1988; Merutka et al., 1993), helix scissoring (separation of the helices), and helix fraying (Rohl & Baldwin, 1994).

Because there is no evidence for slow exchange in the NMR spectra of this peptide, the presence of multiple conformations could lead to essentially three situations. (1) Multiple conformations, such as different loop positions, would give rise to different sets of NOEs that could not be satisfied by single structures. This sort of conformational averaging does not seem to be present because there were no NOE violations in these structures greater than 0.5 Å and few greater than 0.3 Å (Table 2). (2) Conformations could be averaged between those giving NOEs and those with weak or no NOEs, for instance, helix/unfolded equilibrium or scissoring motions of the helices. In this case, the observed intensity of the NOEs is reduced by the fraction of unfolded forms in the ensemble. Because the intensity of the NOEs are also distance ($1/r^6$) dependent, the distance constraints derived tend to be somewhat longer than if a single conformer were present but still biased toward the folded forms of the molecule. In addition, the reduction in intensity of weak NOEs may render them unobservable. This situation leads to the derivation of looser and/or fewer constraints and the calculation of structures with higher RMSDs. There does seem to be evidence for a folded/unfolded equilibrium. The coupling constants in the helical regions of αtα (Fig. 4), although resembling those observed in helical monomeric helices at low temperatures (Osterhout et al., 1989; Bradley et al., 1990; Millhauser et al., 1996), are slightly higher than has been measured for certain proteins (Billeter et al., 1992), suggesting that there is some population of unfolded conformations. Additionally, CD measurements suggest that the helices in the molecule are not fully populated (data not shown). (3) Conformational heterogeneity could give rise to situations where the constraints are looser and multiple derived structures could satisfy the constraints. This seems to be the case in the turn region because all of the derived structures satisfy the NOE constraints and yet the derived structures are notably less precise in this region (Fig. 3), as reflected by the increase in RMSDs (Fig. 4). The increase in coupling constants and chemical shift indices (Fig. 4) in the turn region also supports the notion that this region undergoes more conformational averaging than the interior of the helices.

Because αtα is a designed peptide, it is of primary interest to determine how well the solution structure corresponds to the intended structure. The design of αtα was not computer-aided and no coordinates exist for the design. The target structure for αtα was presented diagrammatically in Fezoui et al. (1994, Fig. 1) and Fezoui et al. (1995a, Fig. 3; note Glu 29 is missing from this figure). Figure 6A is an adaptation of Figure 3 (Fezoui et al., 1995a) and has been corrected to include Glu 29. Figure 6B is a ribbon/space-filling figure of the αtα structure closest to the average. It can be seen from comparison of Figure 6A and B that the overall design goals, namely associating helices connected by a turn region, have been met. However, the two structures differ in several details.

The eight amino acids that comprise the hydrophobic core of the molecule do not show the same packing pattern in the ensemble of NMR-derived structures as in the original design. It is possible that the patterns observed in the NMR structures are a result of the small number of NOEs observed between the helices. The presence of conformational averaging (especially scissors-like opening of the molecule) would lead to slightly larger upper bounds on the distance constraints and fewer observable NOEs. These looser constraints might allow greater variation in the nature of the side-chain interactions and could affect the calculated packing patterns of the amino acids. It should be possible to test these notions experimentally if subsequent molecules can be designed for greater stability.

It is difficult to locate exactly the position of the helices in αtα. αtα was designed to have helices comprising residues 1–17 and 22–38 (Fezoui et al., 1995a). It was realized, however, that the amino acids between the last hydrophobic amino acids and the turn might be less helical than the region encompassed by the hydrophobic interface (Fezoui et al., 1995a). The observed medium- and short-range NOEs suggest that the helices encompass residues 1–17 and 22–38 exactly as designed (Fig. 1). However, the coupling constants and chemical shift indices (Fig. 4A,B) imply that the residues near the turn, namely residues 17, 21, and 22, are less helical. The calculated ensemble of structures (Fig. 3) reflect these differences.

Another aspect of the observed structures that differs from that in the design is the lack of NOE evidence for salt bridges. αtα was designed to have a variety of intra- and interhelical salt bridges. Most of these were positioned to operate across the helices in hopes of gaining stability for helix association (Fezoui et al., 1995a). Because salt bridges are not observed directly in NMR experiments, evidence for their presence would have to come from the observation of side-chain to backbone NOEs (Osterhout et al., 1989), which would serve to orient the side chains in the derived structures. In the ensemble of αtα structures, only possible ion pairs between Arg 6 and Gln 9, and Glu 8 and Arg 26 are observed. The structure of αtα was determined at pH 3.6, which means that the charge on the Glu and Asp residues would be reduced, potentially reducing the strength of any Glu-...X salt bridges.

Although the details of the designed structure and the observed solution structure of αtα differ, it is clear that the overall design goals have been met (Fig. 6A,B). αtα was designed as a pair of interacting helices connected by a turn region, and this structure is achieved in solution. Although the designed and resulting structures were not in perfect correspondence, the molecule is stable enough to allow further development. αtα can now be used to explore the effects of the turn, helix capping interactions, or blocking groups, variations in the hydrophobic interface and details of molecular stability accessed through amide proton exchange.

Materials and methods

Peptide synthesis and purification

αtα was synthesized and purified as described previously (Fezoui et al., 1994). The amino acid composition and molecular mass

were confirmed by amino acid analysis and electrospray mass spectroscopy. Peptide concentrations were determined by triplicate amino acid analysis.

NMR spectroscopy

Samples for NMR spectroscopy were prepared by dissolving lyophilized peptide to a concentration of 3.0 mM in either 90% H_2O/10% D_2O or 99.9% D_2O. The samples contained 20 mM NaCl and the pH was adjusted to 3.6 (uncorrected for isotope effects). All spectra were acquired at 25 °C using a Varian Unity Plus 400 spectrometer in the laboratory of Dr. Gerhard Wagner, Harvard Medical School. A two-dimensional NOESY spectrum (Jeener et al., 1979; Kumar et al., 1980) was recorded with a 200-ms mixing time Z-filtered TOCSY (Rance, 1987) spectra using the DIPSI-2 isotropic mixing sequence (Rucker & Shaka, 1989) were recorded with 60-ms mixing times. These spectra were acquired with 32 scans of 1,024 complex points in t_2 amd 256 complex points in t_1. A DQF-COSY spectrum (Piantini et al., 1982; Rance et al., 1983; Shaka & Freeman, 1983) was acquired with 64 transients of 2,048 complex points in t_2 and 512 complex points in t_1. All spectra were acquired in the phase-sensitive mode using the States–Haberkorn–Ruben method (States et al., 1982).

Data processing was accomplished with the Rowland NMR toolkit (Hoch & Stern, 1993). Spectra were apodized by double-exponential multiplication in t_2 and by phase-shifted sine bell in t_1. The residual water resonance was reduced by filtering of low-frequency components in the time domain (Marion et al., 1989). All spectra were zero filled in both dimensions to yield matrices of 2,048 × 2,048 real points.

Experimental restraints

The program CALIBA (Güntert et al., 1991) was used to calculate upper bound distance restraints from integrated cross peak intensities of a 200-ms mixing time NOESY spectrum using Trp 2 ring interproton distance (HH2 to HZ2 = 2.5 Å) as a calibration distance. Pseudoatom corrections for methyl and methylene protons were applied (Wüthrich et al., 1983) and, in the case of NOEs involving methyl groups, a further correction of 0.5 Å was added to account for the higher apparent intensity of the methyl resonances (Wagner et al., 1987). $^3J_{HN\alpha}$ coupling constants were determined using the program XPEAKFIT (Stern, 1994). t_2 cross sections of the DQF-COSY spectrum were subjected to a nonlinear least-squares fit of the chemical shifts, line widths, and coupling constants. ϕ angles were restricted to the range $(-90°, -40°)$, where the coupling constants were less than 6 Hz. No couplings larger than 8 Hz were found.

Structure calculations

Three-dimensional structures were calculated using a combined distance geometry and simulated annealing approach using the program X-PLOR, version 3.0 (Brünger, 1992). Distance geometry structures were generated using 266 interproton distance constraints and 20 dihedral angle constraints (Havel et al., 1983). The structures were regularized by Powell minimization and refined using a dynamical simulated annealing protocol essentially as described by Nilges et al. (1988). The Powell minimization was performed using the van der Waals, NOE, dihedral, and bond terms for 100 steps followed by another 100 steps, including the angle energy term.

These initial structures were refined by dynamical simulated annealing consisting of 6 ps of dynamics at 2,000 K, followed by a cooling period of 30 ps to a final temperature of 100 K. During the high-temperature dynamics, the VDW force constant was reduced from 20 to 0.003 kcal/M·Å^2 and the improper torsion angle force constant was increased from 0.1 to 1 kcal/M·deg^2. In the cooling stage, the VDW force constant was increased from 0.003 to 4, was held at 200 kcal/M·deg^2, the VDW radius scale factor decreased from 0.9 to 0.75, and the dihedral angle force constant was held at 200 kcal/M·deg^2. These structures were subjected to 200 steps of Powell minimization.

Of 100 initial structures, the 23 lowest-energy structures having no NOE distance constraint violations greater than 0.5 Å or dihedral angle constraint violations greater than 5.0° were analyzed. The coordinates of these structures were deposited with the Protein Data Bank, Brookhaven National Library, Upton, New York 11973, accession code 1abz.

Acknowledgments

We thank the following individuals for support and discussion: Alan Stern, Jeff Hoch, Gerhard Wagner, Eric LaRosa, and Greg Heffron. This work was supported by the Rowland Institute for Science and the National Institutes of Health (GM47467).

References

Adamson JG, Zhou NE, Hodges RS. 1993. Structure, function and application of the coiled-coil protein folding motif. *Curr Opin Biotechnol 4*:428–437.

Baldwin RL. 1989. How does protein folding get started? *Trends Biochem Sci 14*:291–294.

Baldwin RL. 1995. The nature of protein folding pathways: The classical versus the new view. *J Biomol NMR 5*:103–109.

Billeter M, Neri D, Otting G, Qian YQ, Wüthrich K. 1992. Precise vicinal coupling constants $^3J_{HN\alpha}$ in proteins from nonlinear fits of J-modulated [^{15}N,^{1}H]-COSY experiments. *J Biomol NMR 2*:257–274.

Bonvin AMJJ, Brünger AT. 1996. Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? *J Biomol NMR 7*:72–76.

Bradley EK, Thomason JF, Cohen FE, Kosen PA, Kuntz ID. 1990. Studies of synthetic helical peptides using circular dichroism and nuclear magnetic resonance. *J Mol Biol 215*:607–622.

Brünger AT. 1992. *X-PLOR version 3.1 manual*. New Haven, Connecticut: Yale University.

Bundi A, Andreatta R, Rittel W, Wüthrich K. 1976. Conformational studies of the synthetic fragment 1–34 of human parathyroid hormone by NMR techniques. *FEBS Lett 64*:126–129.

Bundi A, Andeatta RH, Wüthrich K. 1978. Characterization of a local structure in the synthetic parathyroid hormone fragment 1–34 by 1H nuclear magnetic resonance techniques. *Eur J Biochem 91*:200–208.

Chakrabartty A, Baldwin RL. 1995. Stability of α-helices. *Adv Protein Chem 46*:141–176.

Ciesla DJ, Gilbert DE, Feigon J. 1991. Secondary structure of the designed peptide Alpha-1 determined by nuclear magnetic resonance spectroscopy. *J Am Chem Soc 113*:3957–3961.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci 4*:561–602.

Dolgikh DA, Abaturov LV, Bolotine IA, Brazhnikov EV, Bychkova VE, Bushliev BN, Gilmanshin RI, Bebedev YO, Semisotnov GV, Tiotopulo EI, Ptitsyn OB. 1985. Compact state of a protein molecule with pronounced small-scale mobility: Bovine alpha-lactalbumin. *Eur Biophys J 13*:109–121.

Dyson HJ, Rance M, Houghten RA, Lerner RA, Wright PE. 1988a. Folding of peptide fragments of proteins in water solution I. Sequence requirements for the formation of a reverse turn. *J Mol Biol 201*:161–200.

Dyson HJ, Rance M, Houghten RA, Lerner RA, Wright PE. 1988b. Folding of peptide fragments of proteins in water solution II. The nascent helix. *J Mol Biol 201*:201–217.

Eccles C, Güntert P, Billeter M, Wüthrich K. 1991. Efficient analysis of protein 2D NMR spectra using the software package EASY. *J Biomol NMR 1*:111–130.

Fezoui Y, Weaver DL, Osterhout JJ. 1994. De novo design and structural characterization of an α-helical hairpin peptide—A model system for the study of protein folding intermediates. *Proc Natl Acad Sci USA 91*:3675–3679.

Fezoui Y, Weaver DL, Osterhout JJ. 1995a. Strategies and rationales for the de novo design of a helical hairpin peptide. *Protein Sci 4*:286–295.

Fezoui Y, Weaver DL, Osterhout JJ. 1995b. De novo design and structural characterization of an α-helical hairpin peptide: A model system for the study of protein folding intermediates. (Correction). *Proc Natl Acad Sci USA 92*:6650.

Goodman EM, Kim PS. 1989. Folding of a peptide corresponding to the α-helix in bovine pancreatic trypsin inhibitor. *Biochemistry 28*:4343–4347.

Goto Y, Aimoto S. 1991. Anion and pH-dependent conformational transition of an amphiphilic polypeptide. *J Mol Biol 218*:387–396.

Güntert P, Braun W, Wüthrich K. 1991. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol 217*:517–530.

Handel TM, Williams SA, DeGrado WF. 1993. Metal ion-dependent modulation of the dynamics of a designed protein. *Science 261*:879–885.

Havel TF, Kuntz ID, Crippen GM. 1983. The theory and practice of distance geometry. *Bull Math Biol 45*:665–720.

Hecht MH, Richardson JS, Richardson DC, Ogden RC. 1990. De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science 249*:884–892.

Hill CP, Anderson DH, Wesson L, DeGrado WF, Eisenberg D. 1990. Crystal structure of α_1: Implications for protein design. *Science 249*:543–546.

Hoch JC, Stern AS. 1993. The Rowland NMR toolkit V3. Cambridge, Massachusetts: Rowland Institute for Science.

Jeener J, Meier BH, Bachmann P, Ernst RR. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J Chem Phys 71*:4546–4553.

Karplus M, Weaver DL. 1976. Protein-folding dynamics. *Nature 260*:404–406.

Karplus M, Weaver DL. 1994. Protein folding dynamics—The diffusion-collision model and experimental data. *Protein Sci 3*:650–668.

Kemp DS. 1990. Peptidomimetics and the template approach to nucleation of β-sheets and α-helices in peptides. *Trends Biotech 8*:249–255.

Kim PS, Baldwin RL. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem 51*:459–489.

Koradi R, Billeter M, Wüthrich K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics 14*:51–55.

Kumar A, Ernst RR, Wüthrich K. 1980. A two-dimensional Overhauser enhancement (2D-NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun 95*:1–6.

Kuroda Y, Nakai T, Ohkubo T. 1994. Solution structure of a de novo helical protein by 2D-NMR spectroscopy. *J Mol Biol 236*:862–868.

Kwon DY, Kim PS. 1994a. Effects of negative charges of a model for bovine pancreatic trypsin inhibitor folding intermediate on the peptide folding. *Biosci Biotechnol Biochem 58*:400–405.

Kwon DY, Kim PS. 1994b. The stabilizing effects of hydrophobic cores on peptide folding of bovine-pancreatic-trypsin-inhibitor folding-intermediate model. *Eur J Biochem 223*:631–636.

Marion D, Ikura M, Bax A. 1989. Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *J Magn Reson 84*:425–430.

Merutka G, Morikis D, Brüschweiler R, Wright PE. 1993. NMR evidence for multiple conformations in a highly helical model peptide. *Biochemistry 32*:13089–13097.

Millhauser GL, Stenland CJ, Bolin KA, van de Ven FJM. 1996. Local helix content in an alanine-rich peptide as determined by the omplete set of $^3J_{HN\alpha}$ coupling constants. *J Biomol NMR 7*:331–334.

Millhauser GL, Stenland CJ, Hanson P, Bolin KA, van de Ven FJM. 1997. Estimating the relative populations of 3_{10}-helix and α-helix in Ala-rich peptides: A hydrogen exchange an high field NMR study. *J Mol Biol 267*:963–974.

Nilges M, Gronenborn AM, Clore GM. 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett 229*:313–324.

Oas TG, Kim PS. 1988. A peptide model of a protein folding intermediate. *Nature 336*:42–48.

Ohgushi M, Wada A. 1983. Molten-globule state: A compact form of globular proteins with mobile side-chains. *FEBS Lett 164*:21–24.

Osterhout JJ Jr, Baldwin RL, York EJ, Stewart JM, Dyson HJ, Wright PE. 1989. ^{1}H NMR studies of the solution conformations of an analogue of the C-peptide of ribonuclease A. *Biochemistry 28*:7059–7064.

Osterhout JJ Jr, Handel T, Na G, Toumadje A, Long RC, Connolly PJ, Hoch JC, Johnson WC Jr, Live D, DeGrado WF. 1992. Characterization of the structural properties of α_1B, a peptide designed to form a four-helix bundle. *J Am Chem Soc 114*:331–337.

Piantini U, Sørensen OW, Ernst RR. 1982. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc 104*:6800–6801.

Quinn TP, Tweedy NB, Williams RW, Richardson JS, Richardson DC. 1994. Betadoublet: De novo design, synthesis, and characterization of a β-sandwich protein. *Proc Natl Acad Sci USA 91*:8747–8751.

Raleigh DP, Betz SF, DeGrado WF. 1995. A de novo designed protein mimics the native state of natural proteins. *J Am Chem Soc 117*:7558–7559.

Raleigh DP, DeGrado WF. 1992. A de novo designed protein shows a thermally induced transition from a native to a molten globule-like state. *J Am Chem Soc 114*:10079–10081.

Rance M. 1987. Improved techniques for homonuclear rotating-frame and isotropic mixing experiments. *J Magn Reson 74*:557–564.

Rance M, Sorensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. 1983. Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun 117*:479–7485.

Rohl CA, Baldwin RL. 1994. Exchange kinetics of individual amide protons in ^{15}N-labeled helical peptides measured by isotope-edited NMR. *Biochemistry 33*:7760–7767.

Rucker SP, Shaka AJ. 1989. Broadband homonuclear crosspolarization 2D NMR using DIPSI-2. *Mol Phys 68*:509–517.

Scholtz JM, Baldwin RL. 1992. The mechanism of α-helix formation by peptides. *Annu Rev Biophys Biomol Struct 21*:95–118.

Shaka AJ, Freeman R. 1983. Simplification of NMR spectra by filtration through multiple-quantum coherence. *J Magn Reson 51*:169–173.

Sieber V, Moe GR. 1996. Interactions contributing to the formation of a β-hairpin-like structure in a small peptide. *Biochemistry 35*:181–188.

States DJ, Haberkorn RA, Ruben DJ. 1982. A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants. *J Mag Res 48*:286–292.

Stern A. 1994. XPEAKFIT. Cambridge, Massachusetts: The Rowland Institute for Science.

Struthers MD, Cheng RP, Imperiali B. 1996. Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science 271*:342–345.

Wagner G, Braun W, Havel T, Schaumann T, Go N, Wüthrich K. 1987. Protein structures by nuclear magnetic resonance and distance geometry: The polypeptide fold of the basic pancreatic inhibitor determined using two different algorithms, DISGEO and DISMAN. *J Mol Biol 196*:611–639.

Wishart DS, Sykes BD, Richards FM. 1992. The chemical shift index—A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*:1647–1651.

Wright PE, Dyson HJ, Lerner RA. 1988. Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding. *Biochemistry 27*:7167–7175.

Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: John Wiley & Sons.

Wüthrich K, Billeter M, Braun W. 1983. Pseudo-structures for the 20 common amino acids for use in the studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. *J Mol Biol 169*:949–961.

Yan Y, Erickson BW. 1994. Engineering of betabellin 14D: Disulfide-induced folding of a β-sheet protein. *Protein Sci 3*:1069–1073.