Structure and Function of an Aromatic Ensemble That Restricts the Dynamics of the Hydrophobic Core of a Designed Helix-Loop-Helix Dimer

Lars Brive, Gunnar T. Dolphin, and Lars Baltzer*

Contribution from the Department of Organic Chemistry, Göteborg University, 412 96 Göteborg, Sweden

Received February 12, 1997[⊗]

Abstract: GTD-43, a polypeptide with 43 amino acid residues, has been designed to fold into a hairpin helix-loop-helix motif that can dimerize to form a four-helix bundle. The mean residue ellipticity is −21000 deg cm^2 dmol^{-1}, and the magnitude is independent of pH in the interval of 2.4−4.6. The chemical shift dispersion of the amide proton and methyl regions of the ^{1}H NMR spectrum is only slightly less than that of Interleukin-4, a naturally occurring four-helix bundle, and much larger than those reported for molten globule structures. The temperature dependence of the ^{1}H NMR spectrum shows that GTD-43 is in slow exchange on the NMR time scale and that the temperature interval for thermal denaturation is narrow. These spectroscopic probes show that GTD-43 exhibits remarkable structural characteristics for a *de novo* designed four-helix bundle, and the structural basis for the observed properties has now been determined. GTD-43 folds into a symmetric dimer, and the secondary structure as well as the formation of the hairpin helix-loop-helix motif has been established from NMR spectroscopy. GTD-43 has been designed to form complementary hydrophobic interfaces between the helices, and an aromatic ensemble consisting of Phe10, Trp13, and Phe34 has been incorporated in the interior of the hydrophobic core to restrict the dynamics and give rise to a well-defined tertiary structure. The interaction between the residues in the aromatic ensemble has been established from the observation of a number of long-range NOEs. The formation of a well-defined tertiary structure in the region of the Phe-Trp-Phe residues has been demonstrated by slow amide proton exchange rates and Trp fluorescence. This is the first report of a designed hydrophobic core in a four-helix bundle based on aromatic residues, and the structural property of GTD-43 suggests that the design principle is of general applicability in the *de novo* design of proteins.

Introduction

Our understanding of protein structure stems from extensive investigations of native proteins and is therefore limited to sequences that fold. The more demanding question of what structure, if any, is formed from any given amino acid sequence is now beginning to be addressed. Non-natural sequences have been designed to form four-helix bundles,[1−6] β-sandwich proteins,[7] and $\beta\beta\alpha$-motifs,[8,9] but the similarities with native proteins have so far been limited, mainly due to the difficulties involved in designing the hydrophobic interior of the folded protein.

In de novo design[10,11] of proteins, the engineering of the hydrophobic core represents a 2-fold problem. Residues must be incorporated that can provide binding energy for folding, and specificity must be encoded into the amino acid sequences so that well-defined tertiary structures are formed. The former problem is readily solved by incorporating a sufficiently large number of hydrophobic residues into the sequence. From inspection of recently reported designed proteins,[11] the emerging rule of thumb for, e.g., four-helix bundles, is to introduce at least one hydrophobic residue per turn of each amphiphilic helix, most frequently a leucine. The latter problem is considerably more complex, and the overwhelming majority of *de novo* designed proteins show the characteristics of molten globules. It was only recently that the first *de novo* designed four-helix bundle, α_2D, was reported to show some of the hallmarks of native proteins, although the structural basis for the observed conformational properties was not reported.[4]

1 5 10 15 20
S-L-Aad-A-Nval-L-Q-E-A-F-R-A-W-L-Q-Y-H-A-A-K-G-

22 25 30 35 40 43
T-G-P-A-Q-D-Q-E-A-L-R-A-F-A-Aad-Q-L-Nval-A-K-I-N

Figure 1. The amino acid sequence of GTD-43. The one-letter code is used for the natural amino acid residues, where A is Ala, D is Asp, E is Glu, F is Phe, G is Gly, H is His, I is Ile, K is Lys, L is Leu, N is Asn, P is Pro, Q is Gln, R is Arg, S is Ser, T is Thr, W is Trp, and Y is Tyr. Aad is α-aminoadipic acid, and Nval is norvaline.

In a recent paper, we reported on the helix-loop-helix dimer GTD-43 (Figure 1), that showed a well-dispersed ^{1}H NMR spectrum with sharp resonances, a narrow temperature interval for thermal denaturation, and a temperature dependence of the ^{1}H NMR spectrum that showed that it was in "slow exchange" on the NMR time scale.[5] It appeared that, for GTD-43, solutions

[⊗] Abstract published in *Advance ACS Abstracts,* August 1, 1997.

(1) Ho, S.; DeGrado, W. *J. Am. Chem. Soc.* **1987**, *109*, 6751−6758.
(2) Olofsson, S.; Johansson, G.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* **1995**, 2047−2056.
(3) Kaumaya, P. T. P.; Berndt, K. D.; Heidorn, D. B.; Trewhella, J.; Kezdy, F. J.; Goldberg, E. *Biochemistry* **1990**, *29*, 13−23.
(4) Raleigh, D. P.; Betz, S.; DeGrado, W. F. *J. Am. Chem. Soc.* **1995**, *117*, 7558−7559.
(5) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 11297−11298.
(6) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955−6962.
(7) Quinn, T. P.; Tweedy, N. B.; Williams, R. W.; Richardson, J. S.; Richardson, D. C. *Proc. Nat. Acad. Sci.* **1994**, *91*, 8747−8751.
(8) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342−345.
(9) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **1996**, *118*, 3073−3081.
(10) Mutter, M.; Vuilleumier, S. *Angew. Chem., Int. Ed. Engl.* **1989**, *28*, 535−676.
(11) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. *Science* **1995**, *270*, 935−941.

S0002-7863(97)00477-0 CCC: $14.00

to both of the design problems discussed above had been found and the basis for its observed structural properties was therefore of considerable interest. In particular, establishing whether the applied design principles had led to the desired results was an important issue as it is a general goal in *de novo* design of proteins.

Strategies for the design of four-helix bundles that focused on the engineering of shape complementary hydrophobic interfaces failed to produce folded proteins with well-defined tertiary structures. It was therefore suggested by Handel *et al.*[12] that complementarity alone was not enough. Some conformational constraint would have to be included that discriminates against alternate conformations and introduces specificity in helix−helix recognition.

The designed conformational constraint in GTD-43 was based on the incorporation of a phenylalanine and a tryptophan in helix I and a phenylalanine in helix II that could interact in the folded state. We now report that we have established, mainly by NMR spectroscopy, that the designed Phe-Phe-Trp ensemble does indeed form a well-defined hydrophobic core and that the induced conformational stability extends over roughly three turns of each helix in the folded helix-loop-helix motif. The design concept therefore appears to be of general applicability in the *de novo* design of proteins.

The reason for the observed conformational stability of α_2D[4] may have been the interplay of two histidines in the hydrophobic core. A monomeric 23 amino acid residue peptide with $\beta\beta\alpha$-topology was also reported recently to have remarkable conformational stability, and at the core of the folded motif were one Tyr, one Phe, and one Leu.[8,9]

Design

GTD-43 was designed to fold into two complementary amphiphilic α-helices connected by a short loop. The choice of amino acid residues was guided by conformational preferences,[13,14] and the amino acid composition was varied extensively to simplify the assignment of the ^{1}H NMR spectrum. The use of artificial amino acids was a natural consequence of this design strategy, and norvaline (Nva)[15] and aminoadipic acid (Aad) were therefore incorporated in the sequence of GTD-43.

In the engineering of the sequence of GTD-43, a generally accepted design strategy for helical structures was used that has been developed over the last decade in several laboratories, as described in a recent review by Bryson *et al.*[11] Capping residues were incorporated to increase α-helical stability (Ser1 and Asn43),[16,17] the helical dipoles were stabilized (Aad3, Lys20, Asp27 and Lys41),[18] and charged residues capable of forming salt bridges were incorporated on the surface of both helices to stabilize the folded versus the unfolded state.[19]

The helices in GTD-43 were joined by a short loop with four residues. Glycines (Gly21 and Gly23) were chosen to terminate the first helix and to provide flexibility next to the conformationally restricted proline (Pro24), which is a good terminator and initiator of α-helices. Threonine (Thr22) was chosen as the second residue in the loop for its helix-destabilizing properties. The loop sequence of GTD-43 is therefore Gly-Thr-Gly-Pro.

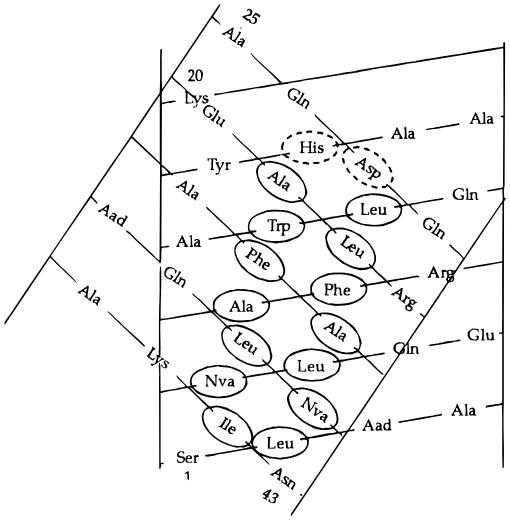

Figure 2. Superimposed helical net diagram of GTD-43 describing the designed contacts between hydrophobic residues with an interhelical angle of approximately −30°. The helices were designed to form a right-handed coiled-coil motif to increase the number of possible hydrophobic contacts between the helices in the folded peptide. Hydrophobic core residues are encircled. His17 and Asp27, the residues in the dashed circles, were designed to form an interhelical ion-pair near the loop.

GTD-43 was designed to fold into a right-handed coiled-coil with an interhelical angle of approximately −30°. In the design of the coiled-coil motif, hydrophobic residues were incorporated into the sequence to form complementary interfaces between the helices. If the sequence is viewed in units of four, *abcd*, *a* and *d* are the hydrophobic residues. If the sequence is viewed in units of eight, *abcdefgh*, the hydrophobic residues are *adeh*. This pattern is different from the more commonly used unit of seven[20] and was not used as a design principle. The octad developed in the iterative design and minimization process was found to provide shape complementarity between coiled helices for an interhelical angle of −30°. The unit of eight is repeated twice in each helix, starting with residues Leu2 and Asp27, except that the residue that corresponds to the last *h* residue in helix I is His and the first a residue in helix II is Asp. The packing of the hydrophobic faces is shown in the "helical net" diagram (Figure 2).[21] Shape complementarity was engineered by placing Ala9, Ala30, and Ala35 opposite to the large aromatic side chains of Phe34, Trp13, and Phe10, respectively. Leucine side chains are smaller than Phe side chains, and leucines are therefore placed opposite leucines and so on. The non-natural amino acid norvaline, a good helix-former,[22,23] was used in the hydrophobic core at positions 5 and 39, and Ile42 was located close to the carboxyl terminal. The variation, again, served to simplify the assignment of the ^{1}H NMR spectrum.

In order to induce conformational stability, the aromatic residues Phe10, Trp13, and Phe34 were incorporated in positions that would allow them to interact in the folded state. This is in

(12) Handel, T. M.; Williams, S. A.; DeGrado, W. F. *Science* **1993**, *261*, 879−885.
(13) Levitt, M. *Biochemistry* **1978**, *17*, 4277−4284.
(14) Richardson, J.; Richardson, D. In *Prediction of protein structure and the principles of protein conformation*; Fasman, G., Ed.; Plenum Press: New York, 1989; pp 1−98.
(15) Abbreviations: Nva, L-norvaline; Aad, L-α-aminoadipic acid; ANS, 8-anilino-1-naphthalenesulfonic acid; CD, circular dichroism; NMR, nuclear magnetic resonance; TOCSY, totally correlated spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; DQF-COSY, double quantum filtered correlated spectroscopy; TFA, trifluoroacetic acid; DCM, dichloromethane; TFE, 2,2,2-trifluoroethanol.
(16) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632−1641.
(17) Serrano, L.; Fersht, A. *Nature* **1989**, *342*, 296−299.
(18) Shoemaker, K.; Kim, P.; York, E.; Stewart, J.; Baldwin, R. *Nature* **1987**, *326*, 563−567.
(19) Merutka, G.; Stellwagen, E. *Biochemistry* **1991**, *30*, 1591−1594.
(20) Hodges, R. S. *Biochem. Cell Biol.* **1996**, *74*, 133−154.
(21) Chothia, C.; Levitt, M.; Richardson, D. *J. Mol. Biol.* **1981**, *145*, 215−250.
(22) Lyu, P.; Sherman, J.; Chen, A.; Kallenbach, N. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5317−5320.
(23) Padmanabhan, S.; Baldwin, R. *J. Mol. Biol.* **1991**, *219*, 135−137.

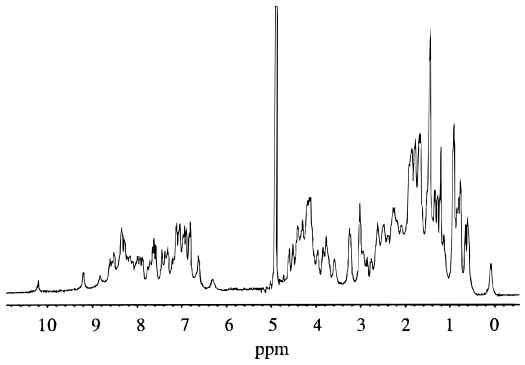

Figure 3. The 500 MHz ^{1}H NMR spectrum of GTD-43 recorded at pH 3.0 in 90% H_2O/10% D_2O at 288 K.

contrast to other four-helix bundles where the hydrophobic cores were composed primarily of leucines, isoleucines, and valines. The underlying assumption was that the interactions between the aromatic side chains would be strong and specific and that conformational rearrangements would be coupled to larger energy barriers than for aliphatic side chains.

Helices form dipoles due to the orientation of the peptide bonds along the helix axis. Calculations have indicated that antiparallel pairing of α-helices is favored over parallel due to the interactions of the dipoles.[24] Although this has little to do with design, the antiparallel helix orientation in a helix-loop-helix motif probably provides some stabilization of the folded motif.

Results

GTD-43 was synthesised using Boc (Boc = *tert*-butoxycarbonyl) chemistry,[25] cleaved from the resin by liquid HF and purified by reversed phase HPLC. GTD-43 was identified by electrospray mass spectrometry (found 4774.6, calcd 4772.4).

The properties of designed proteins are commonly monitored in a qualitative way by recording their ^{1}H NMR and CD spectra and determining their states of aggregation.

The 1D ^{1}H NMR spectrum of GTD-43 in aqueous solution at 288 K and pH 3.0 is well dispersed, and the line widths are narrow (Figure 3). The temperature dependence of the ^{1}H NMR spectrum of GTD-43 in aqueous solution at pH 5.1 and its state of aggregation have been reported previously.[5] The peptide is a dimer in "slow exchange" on the NMR time scale at 288 K.

The CD spectrum of GTD-43 is typical of α-helical proteins[26] with a maximum at 195 nm and minima at 208 and 222 nm. The measured mean residue ellipticity at 222 nm (θ_{222}) of GTD-43 at pH 3.0 was -21000 ± 1000 deg cm^2 dmol^{-1}. Between pH 2.4 and 4.3, θ_{222} was independent of pH, but above pH 4.5, the peptide is poorly soluble as it approaches its isoelectric point at pH 5.9; no reliable values of θ_{222} were obtained. The isoelectric point was determined by isoelectric focusing. The absolute value of the mean residue ellipticity at 222 nm of GTD-43 was also recorded as a function of TFE concentration (Figure 4), and the helical content of GTD-43 increases as a function of TFE as expected.[27] The increase is 15% in the concentration range of 0−40 vol % TFE.

The solution structure was studied by ^{1}H NMR spectroscopy. The resonances were assigned at pH 3.0 and 288 K from the TOCSY, NOESY, DQF-COSY, and HMQC-TOCSY spectra

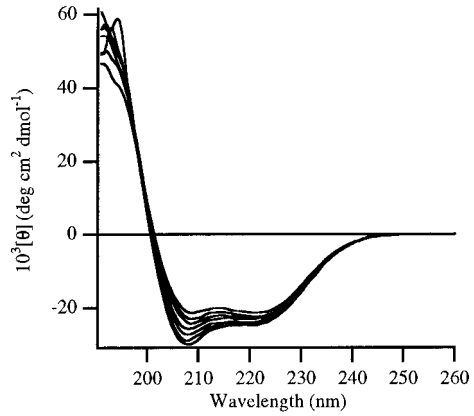

Figure 4. The CD spectrum of GTD-43 as a function of TFE concentration at a peptide concentration of 52 μM in aqueous solution and a pH of 3.0 at 298 K. From top to bottom the TFE concentrations are 0, 5, 10, 15, 20, 30, and 40% (v/v).

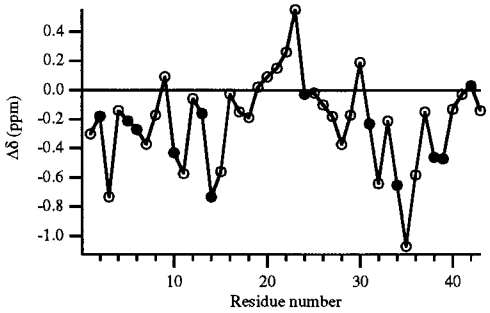

Figure 5. The chemical shift deviation of the α-protons of GTD-43 relative to those of random coils in aqueuos solution at 288 K and pH 3.0. The random coil values for Glu and Leu were used for Aad and Nva, respectively, and the mean value is shown for the α-protons of Gly. Filled circles represent residues in the designed hydrophobic core, including Ala9, Ala30, and Ala35.

(Table 1). Most residue types were identified from the backbone amide proton shifts in the TOCSY spectrum in aqueous solution. Four pairwise overlapping spin systems were resolved after addition of 2% (v/v) TFE, and the amide proton of Leu14 was broadened beyond detection at TFE concentrations less than 8% (v/v). The DQF-COSY spectrum was used to distinguish the single Ile from the leucines and to assign the Phe aromatic protons. The chemical shifts of ring or amide protons of Phe, His, Trp, Gln, and Asn were assigned from their intraresidual connectivity to α- and β-protons. An HMQC-TOCSY spectrum was used for the unambiguous assignment of the ^{15}N-labeled residues Leu38, Ala30, and Ala35. Some stereospecific assignments[28−30] of β-protons were made. The sequential assignments were determined from NH−NH (NN), αH−NH (αN), and βH−NH (βN) NOEs.

The secondary structures of GTD-43 were identified from α-proton chemical shift indices[31] (CSI) and from medium-range NOEs typical of helical structures.[32] The α-proton chemical shifts of residues in helical conformations show negative deviations relative to those of coils, and the deviations were on average −0.3 ppm (Figure 5). The magnitude of the deviations varies with a periodicity of 3−4 residues, which has been

(24) Sheridan, R. P. *Proc. Nat. Acad. Sci.* **1982**, *79*, 4545−4549.
(25) Stewart, J. M.; Young, J. D. *Solid Phase Peptide Synthesis*; Pierce Chemical Co.: Rockford, IL, 1984.
(26) Johnson, W. C. J. *Ann. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 145−166.
(27) Nelson, J.; Kallenbach, N. *Proteins* **1986**, *1*, 211−217.
(28) Hyberts, S. G.; Märki, W.; Wagner, G. *Eur. J. Biochem* **1987**, *164*, 625−635.
(29) Constantine, K. L.; Friedrichs, M. S.; Mueller, L. *J. Magn. Reson.* **1994**, *104*, 62−68.
(30) Cai, M.; Liu, J.; Gong, Y.; Krishnamoorthi, R. *J. Magn. Reson.* **1995**, *107*, 172−178.
(31) Wishart, D. S.; Sykes, B. D. *Methods in Enzymology 239*; Academic Press: San Diego, CA, 1993; pp 363−392.
(32) Wüthrich, K. *Science* **1989**, *243*, 45−50.

Table 1. Proton Resonance Chemical Shift assignments for GTD-43

residue	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	others
Ser-1		4.17	4.43, 4.40		
Leu-2	9.21	4.14	1.74		$C^{\delta}H$ 0.92, 0.85
Aad-3	8.82	3.60	1.93, 1.79	1.65, 1.54	$C^{\delta}H$ 2.34, 2.25
Ala-4	7.72	4.19	1.50		
Nval-5	7.90	4.11	2.01, 1.67	1.39	$C^{\delta}H$ 0.92
Leu-6	8.42	4.05	2.06, 1.76	1.33	$C^{\delta}H$ 0.95, 0.10
Gln-7	8.53	3.96	2.29	2.53	$N^{\epsilon}H$ 7.35, 6.82
Glu-8	8.35	4.16	2.29, 2.23	2.62	
Ala-9	8.18	4.42	1.78		
Phe-10	9.20	4.20	2.95, 2.77		$C^{\delta}H$ 6.31; $C^{\epsilon}H$ 6.98; $C^{\zeta}H$ 7.09
Arg-11	8.61	3.78	2.00, 1.93	1.67	$C^{\delta}H$ 3.25, 3.22; NH 7.46
Ala-12	7.95	4.27	1.67		
Trp-13	8.13	4.50	3.59, 3.55		$C^{\delta 2}H$ 7.23; $N^{\epsilon 1}H$ 10.21
Leu-14	8.13	3.59	1.34	1.28	$C^{\delta}H$ 0.78, 0.66
Gln-15	7.37	3.77	2.07, 1.88	1.97	$N^{\epsilon}H$ 7.32, 6.84
Tyr-16	7.05	4.52	2.40, 2.95		$C^{\delta}H$ 7.13; $C^{\epsilon}H$ 6.82
His-17	7.45	4.45	2.52		$C^{\epsilon 1}H$ 8.01; $C^{\delta 2}H$ 6.93
Ala-18	8.00	4.14	1.29		
Ala-19	8.37	4.35	1.45		
Lys-20	8.60	4.42	1.93, 1.78	1.51, 1,47	$C^{\delta}H$ 1.70; $C^{\epsilon}H$ 3.03; $N^{\zeta}H$ 7.64
Gly-21	8.29	4,14, 4.08			
Thr-22	8.34	4.61	4.35	1.22	
Gly-23	8.35	4.49, 4.53			
Pro-24		4.39	2.41, 2.18	2.09	$C^{\delta}H$ 3.81, 3.72
Ala-25	8.51	4.31	1.48		
Gln-26	8.09	4.23	2.16, 2.11	2.48	$N^{\epsilon}H$ 7.66, 6.95
Asp-27	8.57	4.53	2.93		
Gln-28	8.29	3.96	2.21	2.51	$N^{\epsilon}H$ 7.60, 6.91
Glu-29	7.99	4.16	2.24, 2.29	2.62	
Ala-30	8.25	4.52	1.86		
Leu-31	7.90	4.09	2.01, 1.65	1.40	$C^{\delta}H$ 0.93
Arg-32	7.88	3.71	1.89	1.70, 1.58	$C^{\delta}H$ 3.25; NH 7.39
Ala-33	7.94	4.12	1.81		
Phe-34	8.28	3.98	2.96, 2.59		$C^{\delta}H$ 6.62; $C^{\epsilon}H$ 7.05; $C^{\zeta}H$ 7.13
Ala-35	8.25	3.26	1.35		
Aad-36	8.33	3.75	1.89, 1.82	1.63	$C^{\delta}H$ 2.39, 2.32
Gln-37	7.57	4.18	2.28, 2.17	2.66, 2.48	$N^{\epsilon}H$ 7.43, 7.19
Leu-38	7.78	3.86	1.17	1.11	$C^{\delta}H$ 0.61, 0.57
Nval-39	7.71	3.85	1.71, 1.37	1.28	$C^{\delta}H$ 0.77
Ala-40	7.31	4.20	1.47		
Lys-41	7.77	4.30	1.89	1.51, 1.47	$C^{\delta}H$ 1.70, $C^{\epsilon}H$ 3.03
Ile-42	8.03	4.20	1.88	1.54, 1.21	$C^{\gamma m}H$ 0.94, $C^{\delta}H$ 0.82
Asn-43	8.21	4.60	2.85, 2.77		$N^{\delta}H$ 7.62, 6.90

attributed to the curvature of the coiled-coil structure.[33] Upfield chemical shifts were found in the sequences between residue 1 and 18 and between 26 and 43. Medium-range NOEs, αN(i,i+3), αβ(i,i+3), and αN(i,i+4), were found in the segments from 4 to 17 and from 24 to 42. Two helical segments were thus identified in the sequence.

Downfield chemical shifts that are typically found in β-structures, loops, and turns[31] were observed in the designed loop region from Gly 21 through Gly 23, and in the sequence from residue 18 to 23 only one medium-range NOE was identified. The CSI and the apparent floppiness of the sequence were used to assign the secondary structure of the region as a loop.

The supersecondary structure was identified from long-range NOEs. The formation of the hairpin helix-loop-helix motif was demonstrated by 18 long-range NOE contacts primarily between aromatic side chains and aliphatic protons of residues designed to form the hydrophobic core (Table 2). The head-to-tail geometry of the helices was established from the observed NOEs (Figure 6). The fact that the helices are linked covalently by a loop severely restricts the number of degrees of freedom and the combination of a large helical content, a loop, and a number of long-range NOEs defines the supersecondary structure quite well. The structure of the dimer is elusive since the dimer is symmetric, and only one set of resonances was found in the NMR spectrum. Since the observed long-range NOE contacts can be both inter- and intramolecular in origin, the determination of the dimer structure will have to await isotopic labeling. Crystallographic experiments have so far failed in spite of much effort.

Distance geometry calculations were carried out under the assumption that all of the measured NOEs resulted from intramolecular interactions. A total of 176 NOE restraints were used where 98 were sequential, 60 were medium range, and 18 were long range. The upper and lower limits were set to 5.5 and 1.9 Å, respectively, for all distance restraints. No dihedral angle restraints were used. All of the 50 calculated structures converged into two clusters, one with right-handed and one with left-handed coiled-coils. In 33 of the 50 structures, the NOE violations were less than 0.5 Å. Lowering the limit to 0.2 Å resulted in a set of 22 structures with good geometries with rms deviations from ideal covalent geometry[34] of 0.002 ± 0.0006 Å for bond lengths, 0.42 ± 0.05° for bond angles and 0.24 ± 0.04° for improper torsions. Since there is only one set of NMR proton resonances in the spectrum of GTD-43, only one of the clusters is probable. The average structure within each cluster is a helix-loop-helix motif with a cross-over helical

(33) Zhou, N. E.; Zhu, B.; Sykes, B. D.; Hodges, R. S. *J. Am. Chem. Soc.* **1992**, *114*, 4320−4326.

(34) Brünger, A. T. *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*; Yale University Press: New Haven, CT, 1992.

Table 2. Measured Long-Range NOE Connectivities in GTD-43[a]

name	residue no.	name	residue no.
$C^{\delta}H+$	6	$C^{\epsilon}H+$	34
$C^{\delta}H+$	6	$C^{\zeta}H$	34
$C^{\beta}H+$	9	$C^{\delta 1}H+$	38
$C^{\beta}H+$	9	$C^{\epsilon}H+$	34
$C^{\beta}H+$	9	$C^{\zeta}H$	34
$C^{\alpha}H$	10	$C^{\epsilon}H+$	34
$C^{\alpha}H$	10	$C^{\zeta}H$	34
NH	10	$C^{\epsilon}H+$	34
NH	10	$C^{\zeta}H$	34
$C^{\beta}H+$	13	$C^{\delta}H$	34
$C^{\beta}H+$	13	$C^{\epsilon}H+$	34
$C^{\beta}H+$	13	$C^{\zeta}H$	34
$C^{\delta 2}H$	13	$C^{\alpha}H$	30
$C^{\delta 2}H$	13	$C^{\beta}H+$	33
$C^{\delta 2}H$	13	$C^{\delta}H+$	34
$N^{\epsilon 1}H$	13	$C^{\beta}H+$	33
$N^{\epsilon 1}H$	13	$C^{\epsilon}H+$	34
$N^{\epsilon 1}H$	13	$C^{\zeta}H$	34

[a] For indentification of residue numbers, see Figure 1. Pseudoatoms are denoted by + signs.

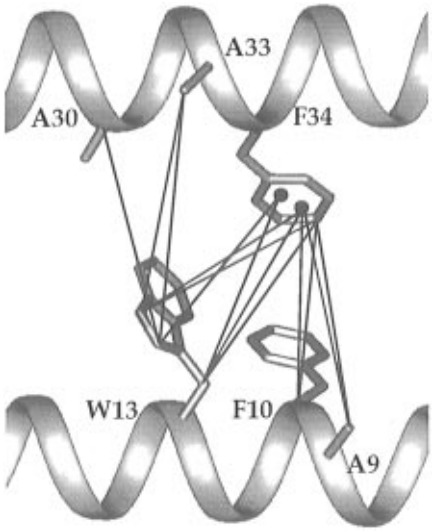

Figure 6. Schematic representation of 13 out of 18 measured long-range ^{1}H NOEs. The lines are drawn between the heavy atoms to which the protons are bound which exhibit NOE contacts. The closed circles in the Phe-34 ring indicate the use of pseudoatoms. The time-weighted average distance between the protons that give rise to an NOE is shorter than 5.5 Å.

angle that approximates 80° (data not shown). However, these structures are preliminary since the NOEs may result from both intra- and intermolecular contacts.

Further information about the molecular environment of Trp13 comes from the fluorescence of the indole residue.[35] The fluorescence maximum of a tryptophan in the apolar environment of a folded protein is blue-shifted relative to that of a tryptophan in a denatured protein and the intensity is increased. The fluorescence spectrum of GTD-43 in the folded state shows a maximum at 344 nm in comparison with the 360 nm measured for the denatured peptide in 5 M GuHCl (Figure 7). The 16 nm blue shift is comparable to what is found in native proteins, and the observed blue shift indicates that Trp13 is buried in the folded structure. Surprisingly, the relative fluorescence intensity of the folded state of GTD-43 is only 20% of that of the unfolded state, suggesting that Trp13 is being quenched.[36−38]

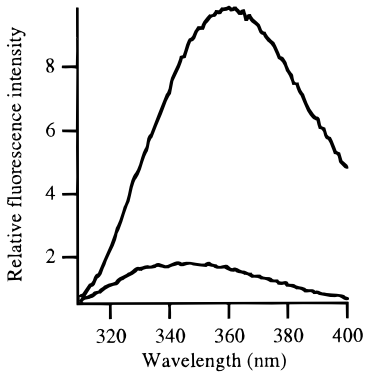

Figure 7. The fluorescence emission spectrum of GTD-43 in aqueous solution (bottom trace) and in 5 M GuHCl (top trace).

The fluorescence of the tryptophan side chain is known to be influenced by imidazole groups, and a possible quencher of Trp13 in GTD-43 is the imidazole moiety of His17 which is located four positions away in the sequence, i.e., one turn away in the α-helix. The pH dependence of the fluorescence spectrum has been determined for GTD-C, a peptide with an identical amino acid sequence to that of GTD-43 except for three residues (His32, Lys36, and Glu3).[39] A sigmoidal increase of the Trp fluorescence was found with a midpoint at 6.5, which is a typical value for the pK_a of histidine.

The thermodynamic stability of a folded polypeptide is an important criterion for the assessment of its similarity with natural proteins. The Gibbs free energy of unfolding of GTD-43 was determined by measurements of θ_{222} as a function of GuHCl concentration at 298 K.[40,41] At a concentration of 2.7 M GuHCl, the helical content was found to be reduced to 50% of that in aqueous solution. Under the assumption that the GuHCl denaturation can be described by a two-state model, folded dimers (F_2) dissociate to form unfolded monomers (U) with a dissociation constant K_U (1).

$$F_2 \overset{K_U}{\rightleftharpoons} 2U \tag{1}$$

The dissociation constant K_U is determined from the total peptide concentration (P_t) and the molar fraction of unfolded peptide (f_U) as in eq 2

$$K_U = \frac{[U]^2}{[F_2]} = 2P_t(f_U{}^2/(1 - f_U)) \tag{2}$$

and f_U is measured from

$$f_U = \frac{\theta_F - \theta_{obsd}}{\theta_F - \theta_U} \tag{3}$$

where θ_{obsd} is the observed mean residue ellipticity at 222 nm for the given GuHCl concentration and θ_F and θ_U are the ellipticities at 222 nm of the folded and unfolded states, respectively. The free energy of unfolding (ΔG_U) is equal to $-RT \ln K_U$. The Gibbs free energy of unfolding in aqueous solution ($\Delta G_U{}^{H_2O}$) was estimated by linear extrapolation of the free energies of unfolding (ΔG_U) at each individual concentration of denaturant to zero GuHCl concentration. The free energy

(35) Lakowicz, Y. *Principles of Fluorecence Spectroscopy*; Plenum Press: New York, 1983.
(36) Shinitzky, M.; Goldman, R. *J. Biochem.* **1967**, *3*, 139−144.
(37) Steiner, R. F.; Kirby, E. P. *J. Phys. Chem.* **1969**, *73*, 4130−4135.
(38) Bushueva, T. L.; Basel, E. P.; Bushueva, V. N.; Burstein, E. A. *Stud. Biophys.* **1974**, *44*, 129−139.
(39) Dolphin, G. T.; Baltzer, L. Unpublished results.
(40) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266−280.
(41) DeFrancesco, R.; Pastore, A.; Vecchio, G.; Cortese, R. *Biochemistry* **1991**, *30*, 143−147.

Figure 8. Schematic representation of short- to medium-range NOEs, signs of α-proton chemical shift deviations from random coil values (chemical shift index, CSI) and proton exchange rates measured at 288 K in aqueous solution. Solid lines indicate observed NOEs, and dashed lines indicate possible NOEs that were unobservable due to spectral overlap. Chemical shift indices and NOEs indicate helical regions between residues 2−18 (CSI), 4−18 (NOE), and 24−43 (CSI), 24−42 (CSI). Circles indicate residues where the amide protons were visible in a NOESY spectrum collected 39 h after addition of D_2O. Residues with slowly exchanging amide protons are centered around the aromatic core residues Phe-10 and Phe-34. Four pairwise overlapping spin systems have been labeled with open circles to show that one or both show slow exchange rates.

of unfolding of GTD-43 in aqueous solution at 298 K was 10.1 ± 0.7 kcal/mol, which corresponds to a dissociation constant of 5×10^{-8} M under the assumption that folding is cooperative.

The conformational stability of *de novo* designed proteins are often evaluated from amide proton exchange rates, thermal denaturation, susceptibility to denaturants such as TFE, and the binding of 8-anilino-1-naphtalenesulfonic acid (ANS).

The amide proton exchange reaction was studied at 288 K and pH* 2.9 (uncorrected) in a series of NOESY experiments in D_2O by monitoring the disappearance of the amide proton intensity as a function of time. Several amide protons showed extended lifetimes, and some protons still had large intensities after 5 days. Amide protons of residues in the sequences from 6 to 14 and from 29 to 39 exhibited slow exchange (Figure 8). Due to the low solubility of the peptide and the demand for short experimental times, the signal-to-noise ratio became too low for accurate cross-peak volume integration. The half-lives were estimated for some protons and were found to be more than 20 h corresponding to protection factors[42] of roughly 100, although this is a conservative estimate since it is based on the part of the resonances that are observable above the noise level.

The absolute value of the mean residue ellipticity, and thus the helical content, of GTD-43 decreases with temperature. At approximately 328 K, a sharp decrease is observed, and at 338 K, as θ_{222} is close to −6000 deg cm^2 dmol^{-1}, the peptide precipitates irreversibly (Figure 9).

TFE reduces rates of amide proton exchange, stabilizes helical structures, and disrupts the tertiary structure of proteins.[43] Molten globule structures have ^{1}H NMR spectra with broadened resonances and poor shift dispersion, and upon addition of TFE, the line widths are reduced and the shift dispersion is decreased. In contrast to what is observed for molten globules, the line widths of GTD-43 were increased upon addition of small amounts of TFE (Figure 10), which is in agreement with the conclusion that GTD-43 is in slow exchange on the NMR time scale. TFE is a denaturant for native proteins, and the effect on the tertiary structure of GTD-43 is probably similar.

ANS is often used to assess the success of *de novo* design of proteins, since ANS fluorescence is strongly enhanced and blue-shifted on binding to the molten globule state.[44] GTD-43 binds ANS, and the emission spectrum of ANS in the presence of GTD-43 has a maximum at 480 nm. However, the determination of the number of binding ANS molecules and the measurement of the binding constant were impossible, because the addition of more than 50 μM ANS to a 50 μM peptide solution causes precipitation.

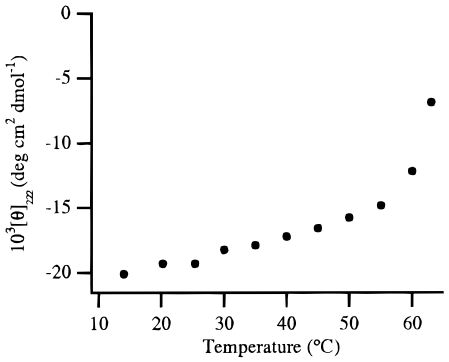

Figure 9. The mean residue ellipticity at 222 nm of GTD-43 as a function of temperature at a peptide concentration of 137 μM in 25 mM KH_2PO_4 at pH 3.5. At 338 K, the peptide precipitates irreversibly.

The ratio of the mean residue ellipticities at 222 and 208 nm ($\theta_{222}/\theta_{208}$) was measured in aqueous solution at pH 3.0 and found to be 0.99, which is commonly interpreted in terms of the formation of a coiled-coil motif.[45,46] An increase in the TFE concentration causes a decrease in the ratio $\theta_{222}/\theta_{208}$ from 0.99 to 0.81 (Figure 11).

Discussion

The structure of GTD-43 has been studied by NMR, CD spectroscopy, and ultracentrifugation. The CD spectrum is typical of helical proteins with a mean residue ellipticity at 222 nm (θ_{222}) of −21000 deg cm^2 dmol^{-1}, which is in the range expected for four-helix bundles.[11] The α-proton chemical shifts (Table 1) and the medium-range NOE connectivities (Figure 8) demonstrate helical conformations in the sequences 2−18

(42) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17*, 75−86.
(43) Olofsson, S.; Baltzer, L. *Folding Des.* **1996**, *1*, 347.
(44) Semisotnov, G.; Rodionova, N.; Razgulyaev, O.; Uversky, V.; Gripas, A.; Gilmanshin, R. *Biopolymers* **1991**, *31*, 119−128.
(45) Cooper, T.; Woody, R. *Biopolymers* **1990**, *30*, 657−676.
(46) Zhou, N.; Zhu, B.; Kay, C.; Hodges, R. *Biopolymers* **1992**, *32*, 419−426.

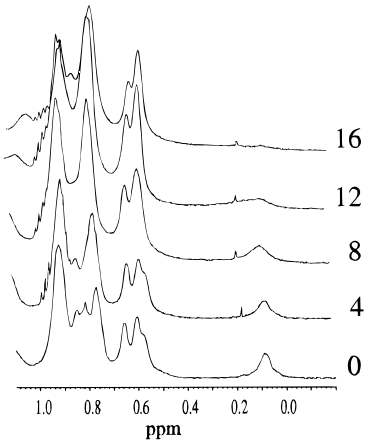

Figure 10. The methyl region of the ^{1}H NMR spectrum of GTD-43 as a function of TFE concentration in aqueous TFE solution at 288 K and pH 3.0. The concentration of TFE is shown as percent of total volume (v/v).

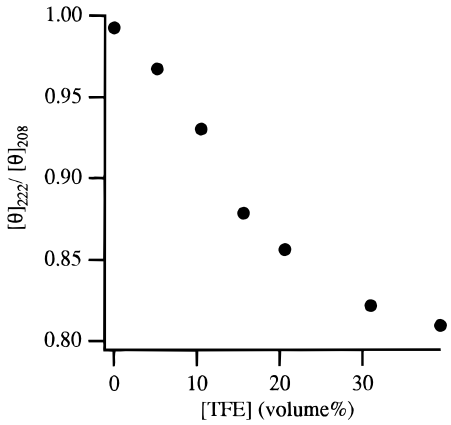

Figure 11. The ratio of mean residue ellipticities at 222 and 208 nm versus TFE concentration at 52 μM peptide concentration in aqueous solution at pH 3.0 and 288 K.

and 24−42. The α-proton chemical shifts and the virtual absence of medium-range NOEs in the sequence 20−23 identifies that segment as a loop. The formation of the hairpin motif is established from long-range "head-to-tail" NOEs (Table 2) with multiple through-space connectivities between Leu 6, Ala 9, Phe 10, and Trp 13 in helix I and Ala 30, Ala 33, Phe 34, and Leu 38 in helix II. Since GTD-43 has been shown to form a dimer,[5] its structure is that of a hairpin helix-loop-helix dimer.

The free energy of unfolding of GTD-43 was 10.1 kcal/mol. Under the assumption that the folding of GTD-43 is a cooperative process, the free energy of unfolding is the same as the free energy of dissociation. The dissociation constant cannot be measured directly but the observed concentration independent CD spectrum[5] in the range of 4−195 μM is compatible with a dissociation constant (K_d) of 5×10^{-8} M, which corresponds to a free energy of dissociation of 10.1 kcal/mol. For a K_d of that magnitude, about 8% of the total peptide concentration is monomeric at 4 μM concentration, and the resulting decrease in mean residue ellipticity is within the limits of error for a measurement of θ_{222} at that concentration. The observation of a concentration independent CD spectrum is, however, incompatible with a substantially larger dissociation constant as the fraction of monomer at 4 μM concentration increases rapidly with K_d, which would be detected in the CD spectrum. Since the free energy of dissociation cannot be smaller than the free energy of unfolding, it may be concluded that both free energies are close to 10.1 kcal/mol and that folding of GTD-43 is cooperative.

Table 3. Proton NMR Chemical Shift Ranges for the *De Novo* Designed Helix-Loop-Helix Dimers GTD-43, α_2D, and SA-42 (43, 35, and 42 residues, respectively) and the Monomeric Native Protein IL-4 (133 residues)[a]

	shift dispersion/ppm			
	NH	C^αH	methyl	ring
GTD-43	2.14	1.35	0.85	0.82
α_2D			0.75	
SA-42[b]	1.12	0.9	0.15[c]	0.18
IL-4	2.63	1.64	0.94	0.82

[a] Values are given for backbone amide protons (NH), α-protons (C^αH), methyl protons (methyl), and aromatic ring protons (ring). [b] In 14% (v/v) TFE. [c] At 0% (v/v) TFE, the value is 0.20 ppm.

The reported free energies of unfolding for the designed helix-loop-helix dimers SA-42 and RA-42 were 12.8 and 8.5 kcal/mol, respectively,[47] and that of α_2B(PRR) was 12.8 kcal/mol.[1] The free energy of dissociation was 7.4 kcal/mol for α_2B[4] and 7.8 kcal/mol for α_3.[3] The thermodynamic stability of GTD-43 with a free energy of unfolding of 10.1 kcal/mol is therefore well within the range of reported values for similar folds. In addition, GTD-43 and α_2D have conformational stabilities that are unique for four-helix bundles. The design principles are, however, very different. The designed conformational constraint of α_2D was composed of nonhydrophobic residues at the hydrophobic−solvent interface, whereas in GTD-43 the Phe-Phe-Trp ensemble is formed within the core as shown by NMR spectroscopy and tryptophan fluorescence.

The ^{1}H NMR chemical shift dispersion, the line widths, and the temperature dependence of the ^{1}H NMR spectrum as well as the narrow temperature interval for thermal denaturation and slow backbone amide proton exchange rates show that GTD-43 has unusual structural properties for a *de novo* designed protein. The ^{1}H NMR chemical shift dispersion compares well with that of other designed four-helix bundles, and it is comparable to that of Interleukin-4 (IL-4),[48] a naturally occurring four-helix bundle with 133 residues (Table 3). The amide protons of GTD-43 cover a shift range of 2.14 ppm, which is 81% of the range covered by IL-4 (2.63 ppm). For comparison, the chemical shift ranges of the molten globule structures SA-42[2] and ALIN[49] are 1.12 ppm for both. The assignment has not been reported for α_2D. The chemical shift dispersion of methyl groups in the hydrophobic core of GTD-43 (0.85 ppm) is comparable to those of IL-4 (0.94 ppm) and α_2D (0.75 ppm), but much larger than those of SA-42 (0.15 ppm) and ALIN (0.18 ppm).

Further chemical shift evidence for the formation of a well-defined tertiary structure comes from the observation of the α-proton of Ala35, which has an upfield shift of 1.1 ppm in comparison to that of a random coil. Ala35 is located in the middle of the hydrophobic core. The chemical shift is probably due to ring current effects of Phe 34, and its magnitude is evidence that its orientation relative to the aromatic ring is restricted over time.

The temperature dependence of the ^{1}H NMR spectrum of proteins is due to conformational changes linked to the transition between folded and unfolded states. In the folded state, the protein populates a small number of conformations. If exchange

(47) Lundh, A.-C.; Broo, K.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* In press.
(48) Powers, R.; Garret, D. S.; March, C. J.; Frieden, E. A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1992**, *31*, 4334−4346.
(49) Kuroda, Y.; Nakai, T.; Ohkubo, T. *J. Mol. Biol.* **1994**, *236*, 862−868.

is slow on the NMR time scale, an increased temperature leads to broadened resonances as the life times of the various conformers become comparable to the inverse of the line widths. If exchange is fast on the NMR time scale, a time-averaged spectrum is observed and the line widths are constant or decrease with increased temperature and increased exchange rates. The latter behavior is typical of molten globules, whereas well-packed proteins show increased line widths and chemical shift changes with increased temperature. The latter is observed for GTD-43,[5] although structural transitions are still observed for some resonances at 288 K. For example, the resonance of one of the methyl groups of Leu38 is strongly broadened and shifted upon changing the temperature from 298 to 288 K. Obviously, the discussion of slow exchange is relevant only to residues that are involved in interactions within the protein, as solvent-exposed side chains rotate rapidly.

TFE is a solvent that is known to reduce amide proton exchange rates, increase the helical content of polypeptides, and disrupt tertiary structures of proteins. The addition of TFE to an aqueous solution of a molten globule reduces the line broadening and shift dispersion since it drives the conformational equilibrium toward helical monomeric species. The effect of adding TFE to an aqueous solution of GTD-43 was the reverse. For small volumes of TFE, line broadening was observed but little decrease in the chemical shift dispersion. Apparently more conformations than the few that are populated in the folded state in aqueous solution are searched as the tertiary structure is disrupted by the less-polar solvent. TFE has been used as a denaturant of proteins to generate long-lived intermediates on the folding pathway,[50,51] and the response of GTD-43 to the change in solvent is probably similar.

The thermal denaturation of GTD-43 shows "in spirit" the behavior of a native protein, although the denaturation is not reversible. A sharp decrease is observed for the absolute value of θ_{222} at 328 K, and at 338 K, as θ_{222} is close to -6000 deg cm^2 dmol^{-1}, the peptide precipitates. Since θ_{222} of a thermally denatured polypeptide is not expected to have a positive value, it can be estimated that at the temperature of precipitation GTD-43 is past the midpoint of denaturation, and the temperature interval for thermal denaturation can therefore be estimated to be narrow and less than 30 °C.

The amide proton exchange rates at 288 K and pH* 2.9 (uncorrected) were slow for several residues (Figure 8). Native proteins with well-defined tertiary structures exchange slower that random coils because stronger hydrogen bonds are formed and accessibility of the solvent is restricted in the tightly packed interior. The rates reflect the dynamic nature of the protein as well as the size. Comparison with exchange rates measured for random coils, the so-called protection factors,[42] provides a measure of how well ordered the structure is. Estimated lower limits of protection factors for GTD-43 were at least 100 for some of the residues. Unfortunately, no protection factors have been measured for IL-4 or α_2D, so no comparisons can be made with proteins of comparable size and fold.

Even more interesting than the protection factors is the distribution of slowly exchanging residues in the sequence (Figure 8). They appear from residue 5 through 14 and from residue 29 through 39 and include the residues for which long-range NOEs were observed. The NOEs and the exchange rates define a region of conformational stability that is extraordinary for a *de novo* designed protein. The loop region, the last turn of helix I, and the first turn of helix II are regions with few medium-range NOEs and fast amide exchange rates. There were no amide protons in slow exchange in the segment from residue 15 to 28, although the chemical shift index indicates helical conformation in helix I as far as 9 residue 17 and in helix II from residue 24. However, the fast exchange rates in helix II are explained by the fact that residues 24 to 27 form the amino end where amide protons are not hydrogen bonded and medium-range NOEs are found that involve residues 24 to 27. The last turn of the designed helix I, however, shows no medium-range NOEs, fast exchange rates and little helical structure. Obviously, the last turn of helix I is less well defined.

The criteria described here demonstrate an exceptional conformational stability of GTD-43 in comparison with those of other designed four-helix bundles. The structurally well-defined region is formed by residues that are centered around the aromatic residues Phe10, Trp13, and Phe34. The other hydrophobic residues in the designed core of GTD-43 are leucines, isoleucines, and norvalines residues with aliphatic side chains that have been used in many designed proteins which formed molten globule structures. The proximity of Phe10, Trp13, and Phe34 in the folded structure is proven by the observation of a number of NOEs (Figure 6, Table 2). The tryptophan fluorescence intensity and wavelength maximum show that the indole residue of Trp is at least partly buried in the hydrophobic interior of the folded helix-loop-helix dimer. The incorporation of the Phe-Phe-Trp is the main structural difference between GTD-43 and other designed proteins, and their proximity in the folded peptide has been proven by NMR spectroscopy. The conclusion that the aromatic residues interact and are responsible for the structural properties of GTD-43 is therefore strongly supported.

The binding of ANS is commonly used as a criterion for discriminating molten globule structures from those of native proteins. GTD-43 does bind ANS, but the structural evidence presented here shows that the tertiary structure of GTD-43 is well defined in the major part of the motif. The ANS binding site of GTD-43 may be the imperfectly packed loop region. ANS binding has been observed for native proteins, too, and its usefulness as a criterion is probably questionable. The information content of ^{1}H NMR spectroscopy is considerably higher concerning the detailed understanding of structure and dynamics, and ANS binding in the case of GTD-43 does not outweigh the combined evidence from NMR and CD spectroscopy that its tertiary structure is well defined.

The design of GTD-43 emphasises shape complementarity of two helices rather than high helix propensity, and the predicted helix propensities are lower than for designed polypeptides with molten globule character. A similar design strategy was proposed by Desjarlais and Handel to lead to more native-like properties in designed proteins on the basis of calculations of backbone optimizations and side chain packing.[52] Less-strong helices adjust better to the geometries imposed by optimum side chain packing. Weak helices may therefore also be important factors that favor the formation of the well-defined tertiary structure of GTD-43.

The α-proton NMR chemical shifts of the residues in the helical segments vary with a periodicity of 3−4 residues (Figure 5), and it has been suggested that periodicity arises as a consequence of helix curvature in the coiled-coil motif.[33] A ratio of the mean residue ellipticities $\theta_{222}/\theta_{208}$ is 1 in aqueous solution, and since it has been argued[45,53] that a ratio that is close to 1 shows that a coiled-coil motif is formed, these two

(50) Sancho, J.; Neira, J. L.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 749−758.
(51) Buck, M.; Radford, R. S.; Dobson, C. M. *Biochemistry* **1993**, *32*, 669−678.
(52) Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1995**, *4*, 2006−2018.
(53) Zhou, N.; Kay, C.; Hodges, R. *J. Biol. Chem.* **1992**, *267*, 2664−2670.

measurements are in agreement. Upon addition of TFE, the ratio decreases to 0.8 as the helix-loop-helix dimer is disrupted by the solvent to form single-stranded helices (Figure 11). TFE is known to disrupt the tertiary structure of proteins and induce helix formation, and GTD-43 is therefore a monomeric helix-loop-helix *nonhairpin* motif in that solvent.

In conclusion, we have designed a helix-loop-helix dimer with exceptional conformational stability for a *de novo* designed protein. The structural basis for its unusual properties is an aromatic Phe-Phe-Trp ensemble that restricts the dynamics in the hydrophobic interior of the folded peptide. The proximity of the residues has been demonstrated by NMR spectroscopy, and the well-defined structural region extends over roughly three turns of each helix in the hairpin helix-loop-helix motif. The use of aromatic residues in designed proteins therefore promises to be a generally applicable strategy for the engineering of motifs with well-defined tertiary structures.

Experimental Section

Peptide Synthesis. GTD-43 was synthesised on a 0.6 mmol scale using a Biosearch 9600 automated peptide synthesiser, Boc chemistry, and a phenylacetamidomethyl (Pam)-linked polymer, with a substitution level of 0.3 mmol/g. Benzyl (Ser, Thr), benzyloxy (Asp, Glu, Aad), ((2-chlorobenzyloxy)carbonyl) (Lys), (benzyloxy)methyl (His), tosyl (Arg), formyl (Trp), and 2,6-dichlorobenzyl (Tyr) groups that are stable in trifluoroacetic acid (TFA) solution were used for side chain protection.

The α-amino protecting groups were removed by 45% TFA in dichloromethane (DCM) containing 2.5% anisole (v/v) as scavenger in a 1 min prewash followed by a 30 min deprotection step, and the free α -ammonium groups were deprotonated by 10% *N*,*N*-diisopropylethyl amine (DIPEA) in DCM. Indole (1%, w/v) was added as an additional scavenger to the TFA solution before incorporation of Trp13.

All couplings were carried out in a mixture of 35% *N*,*N*-dimethylformamide (DMF) and 65% DCM, and two procedures were used. Asn, Arg, Gly, Gln, and amino acids to be coupled to Gln were preactivated for 10 min by 4 equiv of *N*,*N*′-diisopropylcarbodiimide (DIPCDI) and 1-hydroxybenzotriazole (HOBt) and coupled twice for 2 h in a 4-fold excess over peptide. All other amino acids were preactivated for 2 min using 5.5 equiv of DIPCDI and 2.5 equiv of HOBt and reacted for 2h in a 5.5-fold excess over peptide. DIPEA (1.5 equiv) was added to the reaction vessel after 30 min to improve the yield. The residues from Ile42 to Ala30 were coupled for 1h, and the residues from Glu29 to Ser1 were coupled for 2h, except for those that were coupled twice (see above). The Kaiser test was used to monitor the synthesis, and unreacted α-amino groups were acylated by 0.3 M *N*-acetylimidazole in DMF.

The final deprotection and cleavage of the peptide from the resin was accomplished by the low−high HF-method of Tam *et al.*[54] Low HF cleavage was carried out for 1 h at 273 K using 10 mL mixture of hydrogen fluoride, dimethyl sulfide, *p*-cresol, and *p*-thiocresol (25:65:7.5:2.5, v/v) per 1.5 g of peptide resin. High HF cleavage was performed for 1 h at 273 K in a 13 mL mixture of HF, *p*-cresol, and *p*-thiocresol (92.3:5.8:1.9, v/v). The cleaved material was extracted with diethyl ether, ethyl acetate, and DCM to remove the scavengers.

GTD-43 has a sticky quality, and it was found to be difficult to remove the hydrophobic impurities by passing through a Sephadex G25 column, which led to difficulties in reversed phase HPLC purification. To remove the impurities, the peptide was first extracted from the resin with 20% HOAc after which $CHCl_3$ and concentrated HOAc were added until a homogenous mixture was obtained. After titration of the residue with water, the phases separated, and the $CHCl_3$ phase was removed together with the hydrophobic impurities. The aqueous phase was then reduced by evaporation, diluted with water, and lyophilized.

GTD-43 was purified by first passing through a G-25 fine Sephadex column using 0.05% TFA as eluent to remove small terminated peptides and salts. Preparative high-pressure liquid chromatography on a reversed phase Kromasil C-8 column, 25 × 250 mm, was then used with an isocratic mixture of 35% 2-propanol and 0.05% TFA as eleuent and a flow rate of 6 mL min^{-1}. The peptide eluted as a broad peak with a retention time of 16.7 min. GTD-43 was identified by amino acid analysis (BMC, Uppsala, Sweden) and electrospray MS on a VG Analytical AutoSpec mass spectrometer (theoretical 4772.4 found 4774.6). The purity was established from the observation of a single peak in the analytical HPLC chromatogram.

The side chain carboxylic group and the α-amino group of L-α-aminoadipic acid (Aad) was protected for peptide synthesis by first preparing L-aminoadipic acid dibenzyl ester *p*-toluenesulfonate by the method described by Shields *et al.*[55] The diester salt was hydrolyzed by aqueous copper sulfate, and the copper complex of the desired monoester was isolated by the method of Prestige *et al.*[56] The monoester was prepared by the method of Ledger and Stewart[57] by decomposing the copper complex with EDTA. The Boc group was incorporated with a modified procedure using DMSO and di-*tert*-butyl pyrocarbonate giving the desired *N*-Boc-*d*-benzyl-L-aminoadipic acid.[25] A clear oil of Boc-Aad(OBzl) was obtained with 10 mol % EtOAc. Total yield 59%. ^{1}H NMR, $CDCl_3$: δ 1.45 (9H, s, Boc), 1.6−2.0 (4H, m, β- and γ-protons), 2.4 (2H, m, δ-protons), 4.3 (1H, t, α-proton), 5.05 (1H, s, NH), 5.1 (2H, s, benzyl -CH-) 7.35 (5H, m, aromatic benzyl).

The α-amino groups of ^{15}*N*-Leu and ^{15}*N*-Ala were Boc protected by the method described by Stewart and Young using di-*tert*-butyl pyrocarbonate,[25] and the purity was determined by NMR.

NMR Spectroscopy. All spectra were recorded on a Varian UNITY 500 NMR spectrometer operating at 500 MHz equipped with a triple resonance probe (^{1}H, ^{15}N, ^{13}C) and an Ultra Shim system (Resonance Research Inc., Boston). Water suppression was accomplished in all experiments by weak preirradiation of the water signal for 1.5 s. The temperature was 288 K in all experiments except where indicated. Varian's VNMR software was used to process the data, and the analysis was carried out using VNMR and NMRCompass (Molecular Simulations Inc., Burlington) software.

Typical samples were 0.5 mM in peptide in $^1H_2O/^2H_2O$ 90:10, except in the TFE titration and the hydrogen exchange experiments. In the TFE titration experiments, small amounts of TFE were added and the volume was kept constant by removing fractions of the sample after mixing. The hydrogen exchange reaction was monitored after the peptide was dissolved in ^2H_2O. The pH was adjusted by the addition of 0.1 M NaOH.

Typical one-dimensional NMR spectra were the average of 64 transients. Two-dimensional NMR spectra were recorded as data matrices of 2 × 256 increments with 240 transients in each increment and Waltz ^{15}N decoupling. Mixing times of 150 and 50 ms were used in the NOESY and TOCSY spectra, respectively, the 90° ^{1}H pulse was 9 μs, and the spin-lock pulse in the TOCSY experiment was 22 μs. Linear prediction was applied to 512 data points in the indirectly detected dimension, and the spectrum was zero-filled to 4k × 2k data points.

The sample for the NMR proton exchange experiment contained 0.51 mM peptide in 99.80% ^2H_2O (Glaser) which gave a pH* of 2.9 (uncorrected) as measured by a glass electrode. NOESY spectra were collected with 128 increments of 32 scans each. The total experimental time was 6 h. The intrinsic exchange rates were calculated as described by Bai *et al.*[42] with correction for temperature (288 K) and pD (3.3).

Circular Dichroism. Circular dichroism (CD) spectra were acquired on a Jasco J-720 spectropolarimeter in the range of 260−185 nm using quartz cells, 1 nm band width, and 0.25 s response time. The spectra were signal-averaged at least four times, and the solvent base line was recorded separately and subtracted. The spectra were presented and smoothed (binomial, 3 passes) using the software program IGOR Pro (Wavemetrics Inc.). The instrument was routinely calibrated with *d*-10-(+)-camphorsulfonic acid. Samples were prepared by dilution of stock solutions of peptide by volumetric pipets, and the concentrations were determined from the specific absorbance at 280 nm in 6.0 M GuHCl.

(54) Tam, J. P.; Heath, W. F.; Merrifield, R. B. *J. Am. Chem. Soc.* **1983**, *105*, 6442−6455.

(55) Shields, J.; McGregor, W.; Carpenter, F. *J. Am. Chem. Soc.* **1961**, *26*, 1491−1494.

(56) Prestidge, R.; Harding, D.; Battersby, J.; Hancock, W. *J. Org. Chem.* **1975**, *40*, 3287−3288.

(57) Ledger, R.; Stewart, F. *Aust. J. Chem.* **1965**, *18*, 1477−1484.

The temperature dependence of the mean residue ellipticity was determined using a 137 μM aqueous solution of GTD-43 at pH 3.5 in a jacketed quartz cell with a 0.5 mm path length. The temperature was varied in the range from 287 to 339 K and controlled by a Lauda-T water bath. The sample cell temperature was raised in 5 K increments with 10 min delays between spectra to obtain thermal equilibrium.

The CD spectrum of GTD-43 was recorded as a function of TFE using a 52 μM aqueous peptide solution at pH 3.0, and the sample was titrated with TFE from 0 to 40% v/v at 298 K. The pH was not corrected since the effect of TFE on pH is small. The measured mean residue ellipticity was corrected for the dilution of the sample.

A solution of 88 μM GTD-43 was denatured by the addition of 0−5 M GuHCl at 298 K. The samples were prepared from a 530 μM stock solution of peptide by the addition of aliquots from a 7.73 M stock solution of GuHCl (Fluka, Micro Select), followed by dilution with water to the desired volume, and the pH was adjusted to 3.0. The samples were equilibrated for 30 min before analysis. The concentration of the stock solution of GuHCl was determined from its refractive index.[40]

Equilibrium Sedimentation Ultracentrifugation. The equilibrium sedimentation ultracentrifugation was carried out using a MSE Centriscan analytical ultracentrifuge equipped with a photoelectric scanner working at 254 nm. The rotor speed was 29 312 rpm. The partial specific volume was calculated from the amino acid composition. The experimentally determined molecular weight was 8250, and no concentration dependence was detected. A correction due to nonlinearity in the optical system was applied after the instrument had been calibrated by measuring a series of bovine serum albumin solutions of known absorbance. The correction factor was 1/0.92, which gives a molecular weight of 8970. The experiment was carried out in a salt-free solution, which usually gives molecular weights that are too low by 5−10% due to repulsion. The experimentally determined apparent molecular weight therefore corresponds well to the calculated weight of the dimer, which is 9544.

Isoelectric Focusing. Isoelectric focussing was performed using a Pharmacia Immobiline DryPlate with a pH interval of 4.0−7.0 and a Pharmacia Macrodrive 5 power supply. The peptide sample was electrofocused for 3 h using current settings of 3500 V, 1 mA, and 15 W. The electrofocused peptide was visualized by ninhydrin treatment, and the isoelectric point (p*I*) was determined from the linear pH gradient.

Fluorescence. Fluorescence emission spectra were recorded and baseline subtracted on a Spex 1680 Fluorolog τ2 spectrometer at 288 K in a 200 μL cell with a 1 cm path length using a 4 nm band width and a 0.5 s response time. An ANS titration curve was recorded by adding aliquots from an aqueous solution containing 50 μM GTD-43 and 400 μM ANS at pH 3.0 to a solution of 50 μM aqueous GTD-43 at pH 3.0. ANS was excited at 350 nm, and the emission was recorded at 1 nm intervals between 420 and 550 nm. Tryptophan fluorescence spectra were recorded of an aqueous 32 μM peptide solution at pH 3.0 and of the corresponding peptide solution containing 5 M GuHCl as denaturant. The excitation wavelengths were 280, 290, and 295 nm, and the emission spectra was recorded at 1 nm intervals between 310 and 400 nm.

JA970477J