Dissection of the de Novo Designed Peptide αtα: Stability and Properties of the Intact Molecule and Its Constituent Helices†

Youcef Fezoui,‡,§ Emory H. Braswell,‖ Wujing Xian,‡,⊥ and John J. Osterhout*,‡

The Rowland Institute for Science, 100 Edwin H. Land Boulevard, Cambridge, Massachusetts 02142, and The National Analytical Ultracentrifugation Facility at the Biotechnology Center of the University of Connecticut, Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269

Received October 5, 1998; Revised Manuscript Received December 17, 1998

ABSTRACT: αtα is a de novo designed 38-residue peptide [Fezoui et al. (1995) *Protein Sci. 4*, 286−295] that adopts a helical hairpin conformation in solution [Fezoui et al. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3675−3679; Fezoui et al. (1997) *Protein Sci. 6*, 1869−1877]. Since αtα was developed as a model system for protein folding at the stage where secondary structures interact and become mutually stabilizing, it is of interest to investigate the increase in stability that occurs with helix association. αtα was dissected into its component helices and the relative stabilities of the individual helices and the parent molecule were assessed. The ΔG_0 of unfolding of αtα measured by guanidine hydrochloride denaturation was determined to be 3.4 kcal/mol. The equilibrium constant for folding of αtα was estimated from the ΔG_0 as 338 and from hydrogen exchange measurements as 259. The stability of the helices in intact αtα over the individual helices increased by a factor of at least 37 based on amide proton exchange measurements. Sedimentation equilibrium studies showed very little association of the peptides to form either homo- or heterodimers, suggesting that helix association is stabilized by the high effective concentration of the helices caused by the presence of the connecting turn. The effects of salt and pH on the helicity of the component peptides are largely reflected in the intact molecule, implying that short-range interactions still make important contributions to the conformation of the intact molecule even though significant stabilization is caused by helix association.

The Levanthal paradox concerns the disparity in the observed folding time for proteins (10^{-2}−10^2 s) and the time calculated from considering folding as a random process (greater than the age of the universe) (*1*). The inferred consequence of this calculation is that protein folding must be directed in some fashion. This inference has led to the notion that protein folding must proceed through intermediates; thus much of the research in protein folding has been directed toward detection and characterization of folding intermediates (*2*−*4*).

Recent experiments on small proteins in the absence of kinetic blocks have shown that protein folding can proceed to completion on the time scale of milliseconds (*5*). It has been recognized for some time that species accruing in the dead time of stopped-flow refolding experiments contain secondary structure as detected by CD[1] (*6*). Further, amide proton exchange experiments indicate that some folding intermediates may resemble the native structures (*7*). These experiments all indicate that the secondary and tertiary structures of proteins can form quickly and that the intermediates formed seem to resemble native secondary and tertiary structures. Further, the outline of the tertiary structure of proteins may be determined at a stage when the individual elements of secondary structures are only partially stable. So an understanding of the folding pathway must include a detailed understanding of events that happen in the very early stages. This has always been problematic because early folding intermediates are only transiently populated.

One way to address this problem is to employ peptides as model systems for early folding intermediates. Monomeric helices have been extensively studied (*8*, *9*). Several other systems have been developed, some aimed at determining aspects of secondary structure stabilization (*10*−*12*), others at developing simple models for proteins (*13*), a system for investigating salt or solvent dependencies of linked peptides (*14*), systems for studying β-sheets (*15*) and for β-turns (*16*, *17*).

The target of the current study, αtα (Figure 1), was developed as a model system for the study of protein folding at the level of secondary structure association (*18*). αtα was designed de novo to consist of two helices connected by a turn region (*19*). The solution structure of αtα was recently determined by two-dimensional NMR and shows that the molecule contains two helices that associate in solution (*20*).

† This work was supported by the Rowland Institute for Science (Y.F., W.X., and J.J.O.) and by the NSF 9318373 (E.H.B.).

* Corresponding author: Telephone (617) 497-4683; FAX (617) 497-4627; Email osterhout@rowland.org.

‡ The Rowland Institute for Science.

§ 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, MA 02115.

‖ The National Analytical Ultracentrifugation Facility.

⊥ Present address: Hematology/Experimental Medicine Division, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Ave., Boston, MA 02115.

[1] Abbreviations: FMOC, fluoren-9-ylmethoxycarbonyl; CD, circular dichroism; GdnHCl, guanidine hydrochloride.

10.1021/bi9823838 CCC: $18.00
Published on Web 02/11/1999

αtα:
Succinyl-Asp-Trp-Leu-Lys-Ala-Arg-Val-Glu-Gln-Glu-Leu-Gln-Ala-Leu-Glu-Ala-Arg-Gly-Thr-Asp-Ser-Asn-Ala-Glu-Leu-Arg-Ala-Met-Glu-Ala-Lys-Leu-Lys-Ala-Glu-Ile-Gln-Lys-NH_2

α1:
Succinyl-Asp-Trp-Leu-Lys-Ala-Arg-Val-Glu-Gln-Glu-Leu-Gln-Ala-Leu-Glu-Ala-Arg-Gly-NH_2

α2:
Ser-Asn-Ala-Glu-Leu-Arg-Ala-Met-Glu-Ala-Lys-Leu-Lys-Ala-Glu-Ile-Gln-Lys-NH_2

FIGURE 1: Structure of αtα and amino acid sequences of αtα, α1, and α2. (Top) Structural diagram of αtα with the helices represented as ribbons and the side chains of the hydrophobic interface rendered as space-filling models. The drawing represents the NMR-derived structure closest to average structure (*20*). The molecular representation was produced with MOLMOL (*33*). (Bottom) Amino acid sequences of αtα, α1 [the amino-terminal helix, succinyl-(1−18)-NH_2], and α2 [the carboxyl-terminal helix, residues (21−38)-NH_2].

The magnitude of the CD signal of αtα at 25 °C (*18*, *20*) suggests that the stability of αtα is higher than those of peptides corresponding to individual helices of proteins [see ref *8* for a review of some early peptide work and recent papers concerning peptide fragments of myohemerythrin (*21*) and myoglobin (*22*)]. Given that αtα consists of two interacting helices that are mutually stabilizing, it is of interest to know the stabilities of the individual helices and compare this to the stability of helices in the intact molecule.

Another interesting aspect is the degree to which the properties of the individual helices affect their association in the intact molecule. The framework and diffusion−collision−adhesion theories of protein folding predict that the earliest stage of protein folding is the formation of domains or individual elements of secondary structure that interact and are mutually stabilizing (*2*−*4*, *23*, *24*). Presumably these early elements of structure would be of low stability. As they interact and become more stable, the molecule takes on properties of the folded protein. At some point the protein becomes more than a set of secondary structures and takes on properties of its own. It has been demonstrated for short amino acid sequences that the amino acid sequence alone does not determine the structure in the folded protein but instead tertiary interactions are required. The most dramatic demonstration of this is the observation that a 14 residue sequence can take on either a helical or a β-sheet structure depending upon the position in which it is engineered into a protein (*25*). Yet factors that stabilize individual units of secondary structure can have their stabilizing effects carry over into the associated protein. For example, stabilizing interactions built into S-peptide helices carry over into ribonuclease S (*26*). The examination of the properties of the isolated helices of αtα and their comparison to intact αtα provides an opportunity to investigate the balance between the properties of helices and their properties modified by the presence of a hydrophobic interface.

MATERIALS AND METHODS

Peptide Synthesis and Purification. αtα, α1, and α2 (Figure 1) were synthesized by FMOC chemistry on an Applied Biosystems Model 421A peptide synthesizer and purified by reversed-phase high-performance liquid chromatography essentially as previously described (*18*). Amino acid compositions were checked by amino acid analysis and the molecular masses by electrospray mass spectroscopy. Concentrations of αtα and α1 were determined spectrophotometrically (*27*) and that of α2 by triplicate amino acid analysis.

Temperature, pH, Salt, and GdnHCl Unfolding Curves. These experiments were performed by monitoring the CD response at 222 nm with a Jasco model 710 spectropolarimeter and a 1 mm circular cell thermostated to 25 °C. GdnHCl concentrations were determined by refraction (*28*). Temperature unfolding curves were measured at 5 °C intervals by adjusting the temperature manually and allowing the sample to equilibrate for 5 min after the set temperature had been reached before measurement. Peptide concentrations for these experiments were 50 μM. Percent helix was calculated using the baselines derived by Rohl and Baldwin (*29*). The ΔG_0 for the unfolding and m value were obtained from the GdnHCl denaturation data using the method of Santoro and Bolen (*30*).

CD Spectra and Curve Fitting. CD spectra were recorded in 0.2 nm increments from 190 to 300 nm for αtα and α2 and for 195−300 nm for α1. The scan rates were 1 nm/min and the response time was 16 s. The CD spectra were reduced to 195−240 nm in 1 nm increments for analysis. The LINCOMB program (*31*) was used with the Brahms basis set (*32*) for curve fitting.

Calculation of Accessible Surface Area. The solvent-accessible surface area for αtα in each of the NMR-generated structures (*20*) and in the extended form was calculated with MOLMOL (*33*) with a probe diameter of 1.4 Å.

Sedimentation Equilibrium Experiments. Solutions of each of the peptides (α1 and α2) were prepared from solid in 150 mM NaCl and 20 mM phosphate buffer at pH 7.0 to concentrations of 1.0, 0.3, and 0.09 mg/mL. Approximately 300 μL of each solution was loaded into the sample channels of six double-sector cells with an optical path length of 12 mm, resulting in a solution column height of ca. 7 mm. In addition, about 40 μL of a high-density fluorocarbon oil (M&M FC-43) was added in order to raise the bottom of the solution off the cell bottom in order to increase visibility. In total, the solvent channel of each cell was filled with ca. 360 μL of solution. The interference optical system of the Beckman XLI analytical ultracentrifuge was used for measuring the concentration gradients. Sapphire windows were used and the cells were "aged" (*34*) to set the windows. Several water−water blanks were taken at the experimental speed and temperature until the fringe pattern was constant. The cells were centrifuged until the concentration gradients

were at equilibrium at 48K rpm and 4 °C. Scans were made every 3−6 h. Sedimentation equilibrium was determined to have been achieved when successive scans were undetectably different from each other, as determined by a program, MATCH (developed by D. Yphantis), that removes disturbances such as rotor precession. After equilibrium was achieved, the final data sets were taken. Water−water blanks were taken at the experimental speed and were subtracted from the data.

Three mixtures of the peptides were made up to contain a total amount of peptide of about 0.25 mg/mL but in such a way that the approximate molar ratios of α1/α2 were about 1:1, 1:1.3, and 1.4:1. A portion (220 μL) of each mixture was loaded into three separate cells (ca. 40 μL of the fluorocarbon oil was added), producing a solution column of ca. 5−6 mm. Approximately 280 μL of solvent was loaded into the solvent channels of each cell. The cells were then centrifuged to equilibrium at 60K rpm at 4 °C. Scans were taken every 3 h and were analyzed for attainment of equilibrium. Other experimental details are the same for the separate peptides (above). A second set of experiments with mixtures similar to that above, e.g., α1 and α2 at the same molar ratios, were performed at pH 2.5 in 7 mm solution columns centrifuged at 48K rpm at 4 °C.

Data analysis was performed with a nonlinear least-squares program ["Non-Lin" (*35*)]. Global fits can be achieved by combining the data sets taken at different loading concentrations and speeds. The fitting proceeds by one assuming various models with which to fit the data, i.e., the ideal single species model (ISS) and a number of self-association models. For the ISS model, the program produces the value of the *z*-average reduced molecular weight [$M_z' = M_z(1 - \bar{v}\rho)$, where M_z is the true *z*-average molecular weight, $\bar{v}$ is the specific volume, and ρ is the density of the solution] of the molecules and the rms error of fit. If the fit is good, the value represents the molecular weight of the single species present. If the fit is not good, it represents the *z*-average of all molecules present. For self-association models the value of the monomer molecular weight (M_1) and other constants related to the model (i.e., equilibrium constants, second virial coefficient, etc.) and the rms error are produced. The equilibrium constants are presented as ln K_a (where K_a is in the units of $(U/g)^{n-1}$ and n is the degree of association). To convert values of the reduced molecular weight into molecular weight, one needs to know the value of $\bar{v}$ for the molecule in its solvent and the density of the solvent. It is possible, however, to estimate its value from amino acid composition data (*36*). A version of Traube's rules (*37*) was used to estimate the contribution of the succinate group to the value of $\bar{v}$ for the peptides. It was found that the molar volume of this group was less than 5% of the total volume of the molecule. Therefore, although the calculation of the $\bar{v}$ of the succinate residue is only an estimate, any errors in this calculation will affect the value of $\bar{v}$ for the complete molecule only minimally. The values of 0.739, 0.750, and 0.745 mL/g were calculated for α1, α2, and the equimolar combination of the two peptides, respectively. This leads to values of 0.254, 0.243, and 0.249, respectively, for $(1 - \bar{v}\rho)$ since the density of the solvent (from density tables) is about 1.009 g/mL at 20 °C.

Amide Proton Exchange. α1, α2, and αtα were dissolved to 2.0, 3.9, and 3.0 mM, respectively, in D_2O, pH 3.6, containing 20 mM NaCl. One-dimensional ^1H NMR spectra were collected on a Jeol GX400 spectrometer at 18 °C. The amide proton region was integrated to give the total amide proton intensity for each time point.

Amide proton exchange can be described by a general two-step mechanism (*38)* involving a closed species from which amide proton exchange does not occur and an open species that is exchange-competent:

$$\mathrm{C(H)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{O(H)} \overset{k_{int}}{\rightleftharpoons} \mathrm{O(D)} \qquad (1)$$

Here k_{op} is the rate of opening, k_{cl} is the closing rate, and k_{int} is the intrinsic exchange rate from the open species. The observed first-order rate constant for exchange at a particular position, k_{obs}, can be written as (*38*)

$$k_{obs} = (k_{op}k_{int})/(k_{op} + k_{cl} + k_{int}) \qquad (2)$$

Under conditions where the closing rate is faster than the intrinsic exchange rate ($k_{cl} \gg k_{int}$) [this is the EX_2 exchange mechanism (*38*)], the first-order rate constant can be described by

$$k_{obs} = (k_{op}k_{int})/(k_{op} + k_{cl}) \qquad (3)$$

The protection factor, P_f, for an individual amide proton is the ratio by which exchange is slowed in the closed form as compared to the open form and can be defined as

$$P_f = k_{int}/k_{obs} \qquad (4)$$

Equation 3 can be expressed in terms of the protection factor as

$$P_f = (k_{op} + k_{cl})/k_{op} = k_{cl}/k_{op} + 1 \qquad (5)$$

k_{cl}/k_{op} is just the equilibrium constant for the closed form, K_{cl}. Substituting and rearranging leads to an expression for the closed form equilibrium constant as a function of the protection factor:

$$K_{cl} = P_f - 1 \qquad (6)$$

Note that for proteins a second simplifying assumption is usually made, namely, that the protein is mainly present in the closed (folded) form (K_{cl} is large). In this instance, K_{cl} and P_f are approximately equal.

It is found for a number of globular proteins that $K_{fold} = K_{cl}$ for the most slowly exchanging protons (for examples see ref *39*). The interpretation of this correspondence is that the most slowly exchanging protons exchange by a global unfolding mechanism. Protons that exchange more quickly are considered to exchange by partial unfolding mechanisms. In one case where amide protons are observed to exchange more slowly than predicted by K_{fold}, the difference has been attributed to proline isomerization in the unfolded protein (*39*).

The protection factors, P_f, for the slowest exchanging protons of αtα, α1, and α2 were estimated by comparing the slowest decaying components of the exchange data to the average of the intrinsic exchange rates calculated by the method of Bai et al. (*40*). The amide proton intensities of αtα, α1, and α2 follow multiple-exponential decay functions.

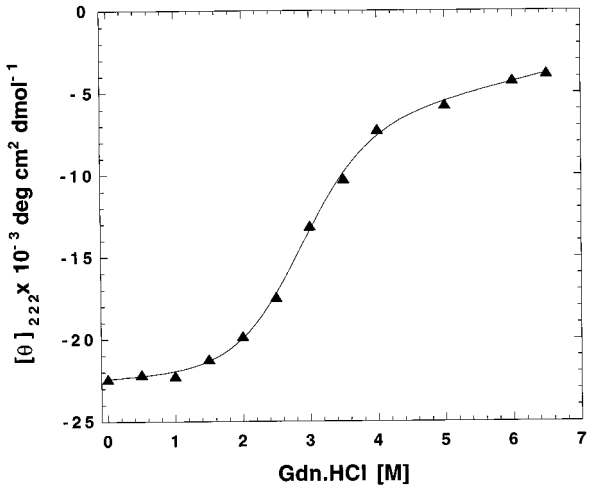

FIGURE 2: GdnHCl unfolding of αtα. The per-residue molar ellipticity of αtα is presented as a function of GdnHCl concentration. Measurements were performed at pH 3.6 and 25 °C. The solid line is the curve fitted to the data by the method of Santoro and Bolen (*30*) as described in Materials and Methods.

For α1 and α2, the slow-decaying components were obtained by fitting the exchange data to double-exponential functions, and for αtα, the exchange data were fitted to three-exponential functions.

RESULTS

GdnHCl Unfolding of αtα. The per-residue molar ellipticity of αtα at 222 nm as a function of GdnHCl is presented in Figure 2. The unfolding transition of αtα in GdnHCl is sigmoidal and exhibits a relatively broad transition. The sigmoidal transition and the presence of folded and unfolded baseline prompted us to analyze these data as a two-state transition. Although the unfolding of a molecule such as αtα is not likely to be two-state, limits on the stability of the molecule can be derived from the analysis. The data were analyzed by the method of Santoro and Bolen (*30*). The ΔG_0 for the unfolding of αtα is 3.4 kcal/mol. The *m* value is 1.22 kcal/(mol·M) and the midpoint of the transition is 2.8 M GdnHCl. The equilibrium constant for folding was calculated from ΔG_0 and was found to be 338.

CD Spectra of αtα, α1, *and* α2. The circular dichroism spectra of the three peptides are presented in Figure 3. All three spectra clearly show helical tendencies. The fraction helix can be calculated from the molar ellipticity at 222 nm on the basis of the molar ellipticity values for unstructured peptides and 100% helix. The helical and unstructured baselines have been estimated in various ways but have been recently derived for short helices as a function of length and temperature by Luo and Baldwin (*41*). Baseline values for peptides of 17 residues were used for α1 and α2. On the basis of these values, the estimates for percentage helix are 34% and 36% for α1 and α2, respectively. On the basis of the GdnHCl denaturation data, the stability of αtα is too high for the baseline method to reflect the helix/coil equilibrium and for this reason a helical estimate is not reported for αtα.

The proportion of various secondary structures in the peptides can also be calculated by fitting the curves to spectral basis sets. As can be seen from Figure 3, all three peptides show minima near 222 nm, characteristic of helix.

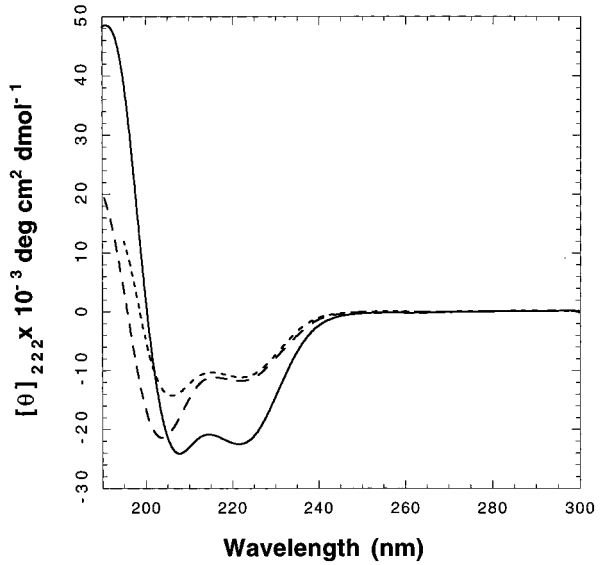

FIGURE 3: CD spectra of the peptides. The per-residue molar ellipticity of α1 (- - -), α2 (— — —), and αtα (—) at pH 3.6 and 25 °C is shown as a function of wavelength.

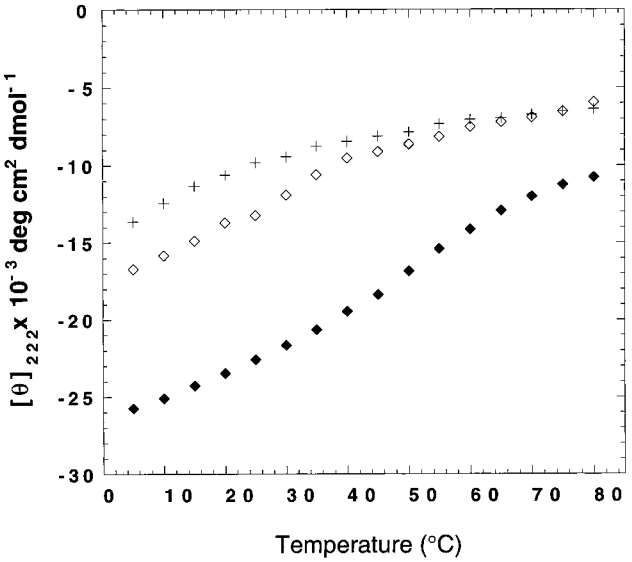

FIGURE 4: Temperature unfolding of the peptides. The per-residue molar ellipticity of α1 (◇), α2 (+), and αtα (◆) is presented as a function of temperature at pH 3.6.

However, they differ substantially in their minima between 200 and 210 nm. The CD spectra were fit with the basis set of Brahms and Brahms (*32*) with the LINCOMB program (*31*). αtα and α2 were optimally fit with only the helix and random basis sets. The percentages of the basis sets used to fit αtα were 78.9% helix and 21.1% coil, and for α2, 43.3% helix and 56.7% coil. α1 required all four of the basis sets for the best fit. The percentages of these were helix, 42.3%; *β*-sheet, 9.9%; *β*-turn, 7.8%; and coil, 40.0%.

Temperature Unfolding of αtα, α1, *and* α2. The changes in molar ellipticity at 222 nm as a function of temperature for the peptides are presented in Figure 4. α1 and α2 seem to come to a common unfolded baseline between approximately 40 and 50 °C. In contrast, αtα appears to exhibit a slightly sigmoidal transition between 40 and 60 °C.

Amide Proton Exchange Experiments. The amide proton exchange results are presented in Figure 5. Also presented is the theoretical exchange curve for the unfolded αtα

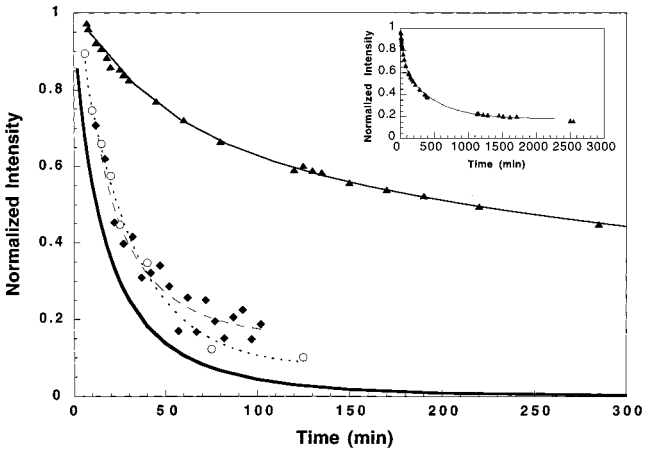

FIGURE 5: Amide proton exchange measurements. Normalized amide proton occupancy as a function of time after dilution into D_2O for αtα (▲), α1 (○), and α2 (◆) at 25 °C and pH 3.6. The lines represent two-exponential (α1 and α2) or three-exponential (αtα) fits to the data. The thick line is the expected exchange from a structureless peptide calculated by the method of Bai et al. (*40*).

calculated by summing the exchange curves of the individual protons using rates calculated according to Bai et al. (*40*). Clearly, the protection of amide protons in α1 and α2 is slightly higher than that predicted for the unfolded peptide. Estimates of the largest protection factors for α1 and α2 were 7.6 and 7.4, respectively. In contrast, the amide protons of αtα exhibit much higher protection factors than the individual peptides. The estimate for the largest protection factor in αtα is 260. The equilibrium constants for folding can be approximated from eq 6 and are found to be 6.6, 6.4, and 259 for α1, α2, and αtα respectively.

Aggregation State of α1, α2, and Mixtures of the Two Peptides. It had been shown previously that αtα was monomeric over a concentration range spanning 5 μM−5 mM as determined by the lack of concentration dependence of NMR and CD signals, by molecular sieve chromatography (*18*), and by sedimentation equilibrium (unpublished results). To assess the aggregation state of α1, α2, and mixtures of the two peptides, equilibrium sedimentation studies were performed and the circular dichroism signal as a function of concentration was examined. Three sets of equilibrium sedimentation data were produced for each peptide (consisting of a total of about ca. 2769 and 2514 data points for α1 and α2, respectively), covering a concentration range from ca. 7 μM to almost 1.5 mM. The data were first fitted globally (with Non-Lin) to the ISS model. This fit produced values of M_z of 2485 ± 24 (Figure 6) and 2336 ± 34 (all molecular weights are reported to 95% confidence limits) as compared with the known values of 2211 and 2131 for α1 and α2, respectively. The rms error associated with these fits was only slightly higher than that associated with the optical system (i.e., 0.018 and 0.019 fringes or about 5 μg/mL), while the observed small systematic error of the residuals indicated a small amount of molecular heterogeneity. The fits were improved marginally (in fact, they are not significantly better as determined by *p*-test) by assuming a monomer−dimer equilibrium. For α1, the M−D model produced a value of M_1 of 2396 ± 34, a value for ln K_a of −3.9 ± 0.3, and a rms error of 0.017 fringes with little systematic error apparent. For α2, this model produced respective values of 2176 ± 34, −3.0 ± 0.3, and 0.018 with no perceptible systematic error. These values of ln K_a would produce about 2% and 5% dimer (by weight) in a 0.4 mM solution of the two peptides, respectively. The somewhat higher values found for M_1 (especially for α1) may be due to a small error in the calculated value of $\bar{v}$, as an error of 1% in the estimate of v can lead to an error of 3% in the value of M_1.

For mixtures of α1 and α2, each set of data (ca. 800 data points each) covering a concentration range from ca. 7 μM to ca. 0.4 mM was fitted with the ISS model. The values of M_z resulting (assuming an average value of $\bar{v}$ for the two peptides of 0.745) were 2244 ± 41, 2046 ± 50, and 2154 ± 41 for the three mixtures, respectively. The rms errors were 0.021, 0.023, and 0.018 fringes with a small amount of heterogeneity observable from the systematic error. No other models returned a significantly better fit with a reasonable value of M_1 (i.e., a value at least as large as the known value for the monomer). These results indicate that there is no appreciable association of the two subunits over the concentration range studied. Similar results were found at pH 2.5.

The ellipticity at 222 nm was examined as a function of concentration and pH for both individual peptides and for mixtures of the two. No significant differences in ellipticity were observed for the isolated peptides or for mixtures of the two over a concentration range of 5 μM−2 mM at pH 2, pH 3.6, and pH 7. These experiments indicate that very little dimerization or aggregation is expected over the concentration range examined in this study.

pH Dependence of Peptide Helicity. The response of the CD signal at 222 nm to pH is shown in Figure 7. All three peptides show striking dependence of helicity on pH. αtα has 15 amino acids with titratable side chains and a succinyl blocking group. The 15 titratable amino acids break down as follows: two Asp, six Glu, four Lys, and three Arg. Each titration curve represents the response of all of these titrating groups. However, all three peptides seem to have approximately biphasic behavior. The pH response curve of each peptide was fit to a two-part Henderson−Hasselbalch equation and the solid lines in the figures represent the fits obtained. The p*K*s derived from the fits were α1, 4.9 and 9.8; α2, 4.6 and 8.6; and αtα, 3.8 and 10.3.

Salt Dependence of Peptide Ellipticity. The dependence of the ellipticity at 222 nm on the concentration of NaCl is presented in Figure 8. The dependence of ellipticity on salt concentration shows four distinct regions for each of the three peptides: an initial decrease in ellipticity from 0 to 200 mM salt, an increase from 200 mM to 1 M, a decrease from 1 M to 2 M, and then a final increase from 2 M onward.

DISCUSSION

A major reason for pursuing these studies was to investigate the stability of the interacting helices in αtα compared to those of the individual helices. This is a difficult problem to address in a simple fashion. Short linear peptides are usually conformational ensembles in solution (*42*, *43*). Estimates of helicity from circular dichroism measurements reflect the average helicity of the entire molecule even though the amount of helix may vary along the length of the peptide due to helix fraying or side-chain interactions. On the other hand, the amide proton exchange technique measures equi-

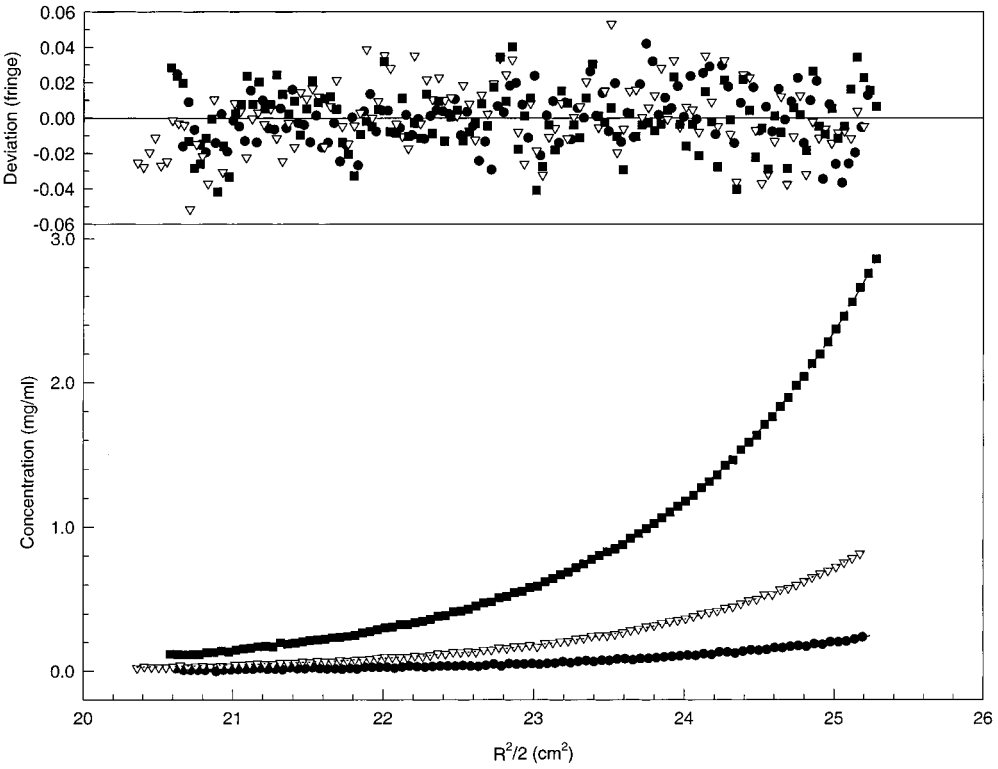

FIGURE 6: Sedimentation equilibrium at 4 °C of α1 at 48K rpm, at 1 (■), 0.3 (▽), and 0.09 mg/mL (●) loading concentration of the peptide. Inset shows residuals for the fit to the ISS model.

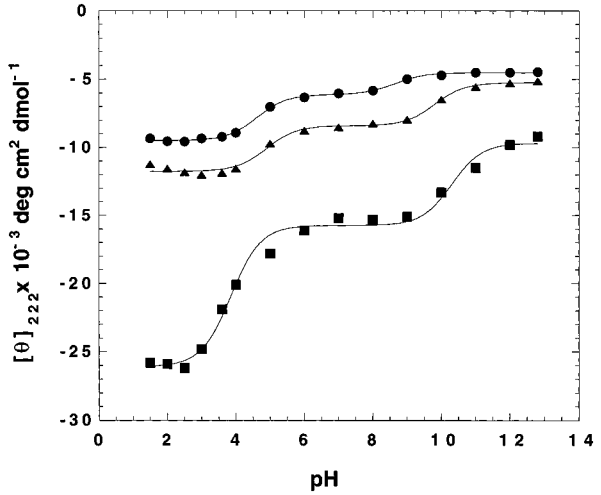

FIGURE 7: pH curves. The per-residue molar ellipticity of α1 (▲), α2 (●), and αtα (■) measured at 25 °C is plotted as a function of pH. Lines represent fits to the data using two-part Henderson−Hasselbalch equations.

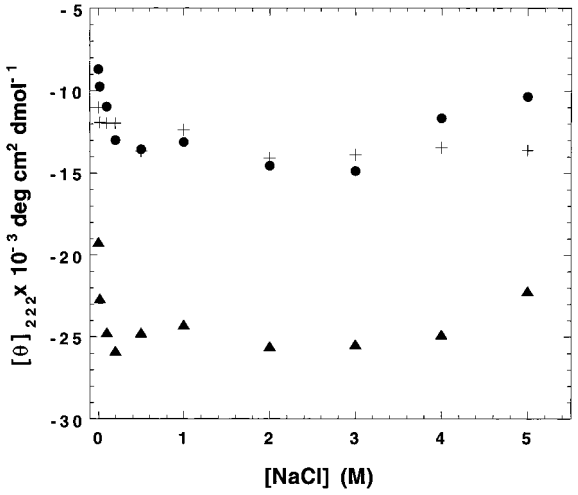

FIGURE 8: Salt dependence of helicity. The per-residue molar ellipticity of α1 (+), α2 (●), and αtα (▲) measured at 25 °C and pH 3.6 is presented as a function of NaCl concentration.

librium constants for closing of the individual amide protons. By taking the group of slowest exchanging protons, the stability estimates are skewed toward positions of highest stability. The estimates obtained from CD measurements for the individual helices, 34% and 35% helix for α1 and α2, respectively, represent equilibrium constants for helix formation of 0.52 and 0.53. The equilibrium constants for folding measured by amide proton exchange are higher, 6.6 and 6.4 [corresponding roughly to 87% and 86% folded at the most stable position(s)]. The accuracy of the CD measurements is likely limited by the suitability of the calculated baselines (*41*). The accuracy of the amide proton exchange measurements is limited by the fitting of a small number of exponentials to a multiexponential process and by the use of the average intrinsic exchange rate to calculate the protection factors.

The equilibrium constant for folding of αtα can be calculated from the GdnHCl unfolding curve, $K_{fold} = 338$, or from the amide proton exchange experiments, $K_{cl} = 259$. Here, as in the peptide experiments discussed above, one experiment (GdnHCl denaturation) measures the overall unfolding while the other (amide proton exchange) measures an equilibrium at one or more of the slowest exchanging sites in the molecule. The two estimates for folding of αtα compare well considering that K_{cl} was calculated from the average intrinsic exchange rate.

The equilibrium constant for folding of the individual helices was approximately 0.5 and 7 by CD and amide proton

exchange. For αtα the estimates are 260 and 340 by amide proton exchange and GdnHCl denaturation. Taking the two measurements derived from amide proton exchange, 7 and 260, implies a relative increase in equilibrium constant of at least a factor of 37 between the isolated and connected helices. The other pair of estimates for the equilibrium constants, 0.5 and 340, yield a much higher upper limit of 680 for the relative increase.

The cooperativity of an unfolding transition can be expressed by the *m* value, which is the slope of the denaturation curve at the midpoint of the transition (measured in kilocalories per mole per molar concentration). Myers (*44*) recently examined *m* values for a large number of globular proteins and these range from a low of 0.58 kcal/(mol·M) (ovomucoid third domain) to a high of 9.7 kcal/(mol·M) (phosphoglycerate kinase). The *m* value of αtα [1.22 kcal/(mol·M), Figure 2] is on the low side of the range, which can be understood in terms of buried surface area of the different molecules. In various treatments of the effects of denaturant molecules on proteins, the *m* value is related to the amount of surface area exposed upon unfolding (*45*, *46*). Myers et al. (*44*) also correlated the *m* values of the proteins to the change in accessible surface area upon unfolding (ΔASA). The relationship that was derived can be used to calculate the *m* value of αtα from its ΔASA. The ΔASA was determined for αtα by subtracting the ASA of the extended form from the average ASA of the ensemble of NMR structures (*20*). The ΔASA for αtα was 1763 Å^2 and the *m* value calculated from the equation of Figure 2A of ref *44* was 1.25 kcal/(mol·M), which is in excellent agreement with the measured *m* value of 1.22 kcal/(mol·M). The correspondence between the calculated and measured *m* values suggests that the low cooperativity of αtα unfolding is expected since the change in the amount of buried surface area is low.

Although the nature of αtα makes it difficult to accurately assign an equilibrium constant to the folding process, it is possible to compare the protection factors for amide proton exchange against a range of peptides and proteins. The range of behavior extends from a protection factor of approximately 1 in unfolded peptides (by definition) to about 10^8-10^9 in the N-terminal helix of cytochrome *c* (*47*). Clearly, αtα is at the lower end of this scale along with peptides, molten globule forms of proteins, and folding intermediates. As examples, the highest protection factor in the low-pH or molten globule form of apomyoglobin is 200 (*48*); in apamin, a 17-residue bee venom peptide, approximately 100 (*49*); in 20-residue alanine-based peptides, 18 (*50*); in the early folding intermediates of T4 lysozyme, 200 (*51*); and in the early intermediates of RNase A, increasing from 4 to 1000 (*52*). This puts αtα in a stability range comparable to folding intermediates, which is consistent with the intent of the original design.

Protection factors can also be compared among designed peptides and proteins. Three examples are α_4, a four-helix bundle protein of 67 residues (*53*), peptide "Velcro", a designed heterodimeric coiled-coil peptide of 60 residues [two monomers of 30 residues each (*54*)] and Model AB, a model αβ motif peptide of 20 residues (*55*). The largest protection factors for these molecules are 10^5 for α_4 (pH 5.0, 25 °C) (*56*), approximately 10^4 for peptide Velcro (pH 4.7, 20 °C) (*54*), approximately 210 for Model AB (pH 4, 5 °C), and approximately 1 for this peptide at 27 °C (*55*). A wide range of exchange behaviors are being observed. The protection factors for the two protein models, α_4 and peptide Velcro, are similar, 10^5 and 10^4, respectively. However, the maximum protection factors estimated from the Δ*G*s of unfolding of these two molecules are quite dissimilar, 10^{16} for α_4 (*56*) and 10^5 for peptide Velcro (*54*). This suggests that while the stabilities of the helices of these molecules are quite similar (they exhibit approximately the same amounts of conformations open to exchange), the exchange from α_4 is dependent upon significant local fluctuations, while peptide Velcro has at least a subset of protons that exchange through global unfolding. The largest protection factors for the two model peptide systems, αtα and Model AB, are 2−3 orders of magnitude smaller than the protein models. These model peptides consist of two elements of interacting secondary structure and have stabilities more in line with folding intermediates than globular proteins.

A significant finding of the present study is that the individual helices do not form significant amounts of either hetero- or homodimers. When αtα was designed it was known that short de novo designed coiled-coils did not associate in solution and that it required a minimum of five heptad repeats for significant dimerization to be observed (*13*). Other studies have shown a varied response to dimerization of helices. Mixtures of coiled-coil peptides will preferentially form heterodimers, likely because of charge interactions (*57*). The antiparallel coiled-coil region of seryl-tRNA synthetase forms a stable hairpin but the individual peptides corresponding to the two helices, absent their connecting turn, do not dimerize (*58*). Designed coiled-coils have also been shown to be stabilized by the presence of a disulfide link between the helices (*59*), although a designed helical hairpin peptide showed no helix association without a stabilizing disulfide bond connecting the central parts of the helices (*60*). These studies show that the stability and specificity of helices are related to the length (which affects the amount of hydrophobic surface area that can be buried upon binding), to charge−charge interactions, and to the presence of connections between the helices. As previously noted, αtα has a relatively small amount of buried surface area, which is likely the major contributing factor to the lack of dimerization of the individual helices.

The lack of dimerization in the absence of the turn region of αtα leads to the idea that the major role of the turn is to increase the effective concentration of the two helices to the point where interaction and mutual stabilization can occur. It may also be involved (passively) in alignment since the first hydrophobic pair of amino acids after the turn (Leu 14 and Leu 25) would have the highest effective concentration. It is not possible from the present data to tell whether the turn also directs the folding in some fashion. These possibilities can be tested with suitable variants of αtα.

A major point of interest is the manner and degree to which the properties of the individual helices affect those of αtα. In the situation where the interaction between the helices is only slightly stabilizing, the properties of the molecule will be balanced between influences arising from within the helices (local interactions) and from those arising from helix association (tertiary interactions). Protein have sufficient stability to allow tertiary interactions to overwhelm local interactions. A case in point is the study of Minor and

Kim (*25*), in which it is shown that an 11 residue segment of amino acids adopts an α-helix when engineered into one portion of a small protein and a *β*-sheet when engineered into another. In this example, the tertiary interactions override the native propensity for conformation of the amino acids. In αtα the amount of energy attributable to the interaction between the helices is small ($K_{fold} = 340$) compared to that available in proteins, where K_{fold} might be 10^5-10^9. For this reason αtα might be expected to exhibit properties of the individual helices. To examine this question the pH and salt dependencies of the three molecules were studied. The expectation was that differences in the response to either salt or pH would indicate the presence of important interactions in the intact molecule that were missing in the individual helices.

The dependence of helicity on pH for the two individual helices is very similar. Both exhibit two arms of the titration curves corresponding to one acid and one base titration. These peptides have five sorts of titrating groups, four arising from amino acid side chains—Glu, Asp, Lys, and Arg—and one arising from the succinyl blocking group. In the individual peptides only one acid and one base titration is observed despite the different kinds of residues. Both the acid and base titrations yield higher helix content as the pH is lowered, with the acid arm exhibiting a larger increase in helicity that the base arm. The increase in helicity in the acid arm of the titration does not seem to be due to protonation of the succinyl group blocking group. The titration observed in αtα is in the wrong direction since succinyl capping interactions or interactions with the helix dipole lead to helix stabilization (*61*). Additionally, removal of the succinyl blocking group from αtα leads to a less helical peptide, as might be expected if the succinyl group were involved in these stabilizing dipole or capping interactions (manuscript in preparation). However, a Glu or Asp residue could be involved in an unfavorable dipole interaction in another part of the chain. Alternatively, the increase in helicity in the acid arm of the titration could arise from changes in the *s* value for helix formation of Glu.

All three molecules exhibit both acid and base titrations at approximately the same p*K*s. The absence of significant changes in the positions of acid and base arms of the titration suggests that no strong charge−charge interactions occur between the helices in the intact molecule. The absence of change in the p*K*s also suggests that the structure of αtα is strongly influenced by structure in the individual helices from which the pH effects largely arise. The overall structure results from a balance of effects operating on the helices and on the association of the helices through the hydrophobic interface.

That factors which have an effect on the individual helices also strongly affect the structure of αtα is further supported by the response to salt (Figure 8). All three peptides have essentially the same shape for the salt curves. Each curve has four parts. Between 0 and 100 mM NaCl helicity increases. This has been attributed to the neutralization of an unfavorable helix dipole interaction (*61*). In the 0.1−3 M salt range all three peptides show two different salt behaviors, the first of which (0.1−1 M) is helix-destabilizing and the second of which (1−3 M) is helix-stabilizing. These concentrations of salt usually effect salt bridges or electrostatic interactions (*62*). The early decrease in helicity must correspond to the neutralization of favorable salt bridges, while the later increase in helicity must correspond to the neutralization of stronger, unfavorable electrostatic interactions. Finally, in the range between 3 and 5 M salt a general decrease in helicity is observed. This has been attributed to Hofmeister salt effects (*62*). Since the responses to salt of all three peptides are similar, the electrostatic interactions that affect helicity in the individual peptides are still largely operative in αtα.

During the process of protein folding, the earliest intermediates might be very local—just stretches of a few residues stabilized by local side-chain interactions. These sorts of intermediates would have time to search quite a large part of the available conformational space [this was a suggestion made in the diffusion−collision−adhesion model of protein folding (*23*)]. As these small, very local elements of structures enlarged, the properties important to stabilization of helices would become operative. These might be charge−charge interactions, helix−dipole interactions, helix propensities, and capping interactions. As individual helices interact through hydrophobic interfaces, different sorts of stabilizing influences could become important. The hydrophobic interface is certainly one and the presence and nature of the turn residues is another. The properties of these ensembles would be different from the properties of the individual helices because of the association. Finally, as the native structure of the molecule is achieved, the properties of the ensemble might change again as close packing becomes important and as charge residues are drawn together on the surface of the molecule.

The purpose of developing αtα, a model for an early folding intermediate, was to provide a system where the properties of interacting helices could be studied in a structure with a stability approximating that of early folding intermediates. The questions of importance are how stable is the molecule, how much stability is developed in the ensemble as a result of helix association, how much do the different helix-stabilizing factors contribute to stabilization in the individual helices and in intact αtα, and what is the balance between properties of individual helices and properties of the ensemble? The present report is a start toward answering these questions. The picture of the molecule that emerges is one in which a fair degree of stability is imparted to the individual helices as a result of being in the intact molecule (about 37-fold) and yet in which the charge interactions that stabilize the individual helices are largely carried over into the ensemble. The model system can now be used to test in detail the effects of the hydrophobic interface, the contribution of the turn, and the detailed effects of charge interactions. As a beginning, the point by point stability of the molecule is being examined by partial ^{15}N labeling and hydrogen exchange studies.

Finally, this study has a fundamental implication for protein folding: the tertiary structure is attained at a relatively low stability. The ease with which tertiary structures can be designed by considering the hydrophobic/hydrophilic pattern as is done with minimalist design (*53*) or the high proportion of folded molecules found using genes designed to give a predetermined pattern of polar and nonpolar amino acids (*63*) imply that code for the overall shape of the molecule is relatively simple. The rapid formation of intermediates suggests that the overall shape can be attained quickly during

folding (*6*, *7*). The continued use of αtα as a model system for the early stabilization of tertiary structure should allow the balance of influences among the turn, hydrophobic interface, electrostatic interactions, intrinsic helix-forming propensities, and others to be studied systematically.

ACKNOWLEDGMENT

We thank Dr. Diane Schaak and Dr. Peter Connolly for critical reading of the manuscript.

REFERENCES

1. Levinthal, C. (1968) *J. Chim. Phys. 65*, 44−45.
2. Kim, P. S., and Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459−489.
3. Baldwin, R. L. (1995) *J. Biomol. NMR 5*, 103−109.
4. Karplus, M., and Weaver, D. L. (1994) *Protein Sci. 3*, 650−668.
5. Baldwin, R. L. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2627−2628.
6. Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87−103.
7. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) *Nat. Struct. Biol. 3*, 782−787.
8. Scholtz, J. M., and Baldwin, R. L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 95−118.
9. Chakrabartty, A., and Baldwin, R. L. (1995) *Adv. Protein Chem. 46*, 141−176.
10. Oas, T. G., and Kim, P. S. (1988) *Nature 336*, 42−48.
11. Kwon, D. Y., and Kim, P. S. (1994) *Biosci. Biotechnol. Biochem. 58*, 400−405.
12. Kwon, D. Y., and Kim, P. S. (1994) *Eur. J. Biochem. 223*, 631−636.
13. Lau, S. Y. M., Taneja, A. K., and Hodges, R. S. (1984) *J. Biol. Chem. 259*, 13253−13261.
14. Goto, Y., and Aimoto, S. (1991) *J. Mol. Biol. 218*, 387−396.
15. Kemp, D. S. (1990) *Trends Biotechnol. 8*, 249−255.
16. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., and Wright, P. E. (1988) *J. Mol. Biol. 201*, 161−200.
17. Stradley, S. J., Rizo, J., Bruch, M. D., Stroup, A. N., and Gierasch, L. M. (1990) *Biopolymers 29*, 263−287.
18. Fezoui, Y., Weaver, D. L., and Osterhout, J. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3675−3679.
19. Fezoui, Y., Weaver, D. L., and Osterhout, J. J. (1995) *Protein Sci. 4*, 286−295.
20. Fezoui, Y., Connolly, P. J., and Osterhout, J. J. (1997) *Protein Sci. 6*, 1869−1877.
21. Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol. 226*, 795−817.
22. Reymond, M. T., Merutka, G., Dyson, H. J., and Wright, P. E. (1997) *Protein Sci. 6*, 706−716.
23. Karplus, M., and Weaver, D. L. (1976) *Nature 260*, 404−406.
24. Baldwin, R. L. (1989) *Trends Biochem. Sci. 14*, 291−294.
25. Minor, D. L., Jr., and Kim, P. S. (1996) *Nature 380*, 730−734.
26. Mitchinson, C., and Baldwin, R. L. (1986) *Proteins: Struct., Funct., Genet. 1*, 23.
27. Elwell, M. L., and Schellman, J. A. (1977) *Biochim. Biophys. Acta 494*, 367−383.
28. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
29. Rohl, C. A., and Baldwin, R. L. (1997) *Biochemistry 36*, 8435−8442.
30. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry 27*, 8063−8065.
31. Perczel, A., Park, K., and Fasman, G. D. (1992) *Anal. Biochem. 203*, 83−93.
32. Brahms, S., and Brahms, J. (1980) *J. Mol. Biol. 138*, 149−178.
33. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
34. Ansevin, A. T., Roark, D. E., and Yphantis, D. A. (1970) *Anal. Biochem. 34*, 237−261.
35. Johnson, M. L., Correia, J. J., Halvorson, H. R., and Yphantis, D. A. (1981) *Biophys. J. 36*, 575−588.
36. Cohn, E. J., and Edsall, J. T., Eds. (1943) *Proteins, Amino Acids, and Peptides*, pp 370−381, Reinhold Publishing Corp., New York.
37. Cabani, S., Gianni, P., Molica, V., and Lepori, L. (1981) *J. Solution Chem. 10*, 563.
38. Hvidt, A., and Nielson, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
39. Mullins, L. S., Pace, C. N., and Raushel, F. M. (1997) *Protein Sci. 6*, 1387−1395.
40. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
41. Luo, P., and Baldwin, R. L. (1997) *Biochemistry 36*, 8413−8421.
42. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) *Biochemistry 27*, 7167−7175.
43. Osterhout, J. J., Jr., Baldwin, R. L., York, E. J., Stewart, J. M., Dyson, H. J., and Wright, P. E. (1989) *Biochemistry 28*, 7059−7064.
44. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
45. Schellman, J. A. (1978) *Biopolymers 17*, 1305−1322.
46. Alonso, D. O. V., and Dill, K. A. (1991) *Biochemistry 30*, 5974−5985.
47. Wand, A. J., Roder, H., and Englander, S. W. (1986) *Biochemistry 25*, 1107−1114.
48. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) *Science 249*, 1544−1548.
49. Dempsey, C. E. (1986) *Biochemistry 25*, 3904−3911.
50. Rohl, C. A., and Baldwin, R. L. (1994) *Biochemistry 33*, 7760−7767.
51. Lu, J. R., and Dahlquist, F. W. (1992) *Biochemistry 31*, 4749−4756.
52. Udgaonkar, J. B., and Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197−8201.
53. Regan, L., and DeGrado, W. F. (1988) *Science 241*, 976−979.
54. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) *Curr. Biol. 3*, 658−667.
55. Butcher, D. J., Bruch, M. D., and Moe, G. R. (1995) *Biopolymers 36*, 109−120.
56. DeGrado, W. F., Raleigh, D. P., and Handel, T. (1991) *Curr. Opin. Struct. Biol. 1*, 984−993.
57. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992) *Cell 68*, 699−708.
58. Oakley, M. G., and Kim, P. S. (1997) *Biochemistry 36*, 2544−2549.
59. Hodges, R. S., Zhou, N. E., Kay, C. M., and Semchuk, P. D. (1990) *Pept. Res. 3*, 123−137.
60. Kuroda, Y., Nakai, T., and Ohkubo, T. (1994) *J. Mol. Biol. 236*, 862−868.
61. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., and Baldwin, R. L. (1987) *Nature 326*, 563−567.
62. Fairman, R., Shoemaker, K. R., York, E. J., Stewart, J. M., and Baldwin, R. L. (1990) *Biophys. Chem. 37*, 107−109.
63. Kamtekar, S., Schiffer, J. M., Xiong, H. Y., Babik, J. M., and Hecht, M. H. (1993) *Science 262*, 1680−1685.

BI9823838