# Mutagenic and Thermodynamic Analyses of Residual Structure in the α Subunit of Tryptophan Synthase†

Gloria Saab-Rincón,‡ Peter J. Gualfetti, and C. Robert Matthews*

*Department of Chemistry and Center for Biomolecular Structure and Function, The Pennsylvania State University, University Park, Pennsylvania 16802*

*Received July 26, 1995; Revised Manuscript Received December 11, 1995*⊗

ABSTRACT: The α subunit of tryptophan synthase from *Escherichia coli* has been previously shown to contain residual structure at 5 M urea, conditions where the secondary structure is entirely disrupted and the tyrosine residues are exposed to solvent [Saab-Rincón, G., Froebe, C. L., & Matthews, C. R. (1993) *Biochemistry 32*, 13981−13990]. The residual structure can be monitored by one-dimensional NMR spectroscopy studies of histidine 92 whose Cϵ proton is sensitive to the slow exchange between this form and the unfolded protein. The temperature dependence of the cooperative urea-induced unfolding transition between intermediate and unfolded forms demonstrates that this process involves negative values for both the enthalpy and entropy changes at 25 °C. The effects of replacements of several nonpolar side chains adjacent to histidine 92 on the slopes and midpoints of the unfolding transition curve show that these side chains participate in the residual structure. A 15-residue peptide spanning histidine 92 and the mutated residues, however, is not sufficient to define this structure. These results demonstrate that the residual structure in the α subunit is stabilized by the hydrophobic effect and may involve side chains which are distant in sequence to histidine 92.

The accurate assessment of equilibrium and kinetic protein-folding mechanisms requires an understanding of the structures of the two end states in this complex conformational transition. X-ray crystallography and NMR[1] spectroscopy provide the necessary information on the native, functional form. The dynamic nature of the exceedingly large ensemble of conformations which define the unfolded state makes its characterization much more difficult. This problem is compounded by the dependence of the structure of the unfolded state on the environmental conditions (Dill & Shortle, 1991; Matthews, 1993). Chemical denaturants such as urea and guanidine hydrochloride appear to be most effective in disrupting secondary and tertiary structure in proteins (Tanford, 1968). Residual structure persists, however, at high or low temperatures and either alkaline or acidic pH for a variety of proteins (Arcus *et al*., 1994; Buck *et al*., 1994; Griko *et al*., 1994a,b; Jacobs & Fox, 1994).

Although thermodynamic and kinetic analyses are simplified if the unfolded state is a random coil, it is difficult to demonstrate how closely a given unfolded protein might resemble a random coil. Viscosity (Reynolds & Tanford, 1970) and size exclusion chromatography (Shalongo *et al*., 1993; Uversky & Ptitsyn, 1994) offer direct measures of the radius of an unfolded protein, a parameter which can be compared to that predicted for a random coil by statistical mechanical theory (Flory & Miller, 1966). However, the simplifying assumptions in this theory and the global nature of these measurements make it possible to miss localized elements of residual structure. In practice, an unfolded protein is often assumed to be a random coil if the native secondary and tertiary structure detected by absorbance, fluorescence, and circular dichroism spectroscopy are completely dissipated at high denaturant concentrations.

The potential limitations of these optical methods, however, have recently been highlighted by the detection of residual structure in several proteins which appeared to be fully unfolded by these criteria. Multidimensional NMR techniques were used to determine the structure of a well-defined hydrophobic cluster in urea-denatured 434 repressor (Neri *et al*., 1992a). This cluster was found to involve nonpolar side chains, Val54, Val56, Trp58, and Leu59, that are constrained to be in spatial proximity by their proximity in sequence. Fluorescence energy transfer between a tryptophan donor and a chemically introduced acceptor in staphylococcal nuclease indicated that the distance between these sites continues to increase at guanidine hydrochloride concentrations where the tryptophan is exposed to solvent and the secondary structure is disrupted (James *et al*., 1992). $^{19}$F-labeled Trp82 in rat intestinal fatty acid binding protein displays chemical shift effects which continue after the secondary structure is lost; this residue only becomes completely exposed to solvent above 5 M urea (Ropson & Frieden, 1992).

A particularly intriguing example of residual structure is provided by the α subunit of tryptophan synthase from *Escherichia coli*. The α subunit has been found to contain nonrandom structure near one of its histidines at 5 M urea, a concentration of denaturant that is sufficient to disrupt the secondary structure and expose the tyrosines to solvent (Saab-

† This work was supported by the National Institutes of Health through Grant GM 23303 to C.R.M.

* Author to whom correspondence should be addressed.

‡ Present address: Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, Cuernavaca, Morelos México.

⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1996.

[1] Abbreviations: AB, absorbance; CD, circular dichroism; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; DTE, dithioerythritol; $K_2$EDTA, ethylenediaminetetraacetic acid, dipotassium salt; $NaDodSO_4$, sodium dodecyl sulfate; NMR, nuclear magnetic resonance; UV, ultraviolet; WT, wild type.

0006-2960/96/0435-1988$12.00/0 © 1996 American Chemical Society

Rincón *et al*., 1993). One-dimensional (1D) proton NMR methods were used to show that a stable intermediate (designated I2), possessing residual structure near His92, cooperatively unfolds when the urea concentration is increased from 5 to 7.5 M at 25 °C. Neither the identity of other side chains involved in the residual structure nor the forces which stabilize I2 and result in a cooperative, urea-induced unfolding transition were resolved in that study.

This paper reports the results of mutation of six amino acids near His 92 in the α subunit. Perturbations in the equilibrium unfolding profiles defined by the His92 Cϵ proton demonstrate that five of the six altered side chains participate in the residual structure. The absence of this behavior in a 15-reside peptide spanning His92 and those mutated sites suggests that this intermediate may involve nonlocal interactions. A thermodynamic analysis shows that the hydrophobic effect is primarily responsible for the stability of the intermediate.

## MATERIALS AND METHODS

*Chemicals and Reagents.* Ultrapure urea was purchased from ICN Biomedicals (Costa Mesa, CA) and used without further purification. $^2H_2O$ (100% pure)and $^2HCl$ (99.0% pure) were purchased from MSD Isotopes (Montreal, Canada); $NaO^2H$ (99.0% pure) was purchased from ICN Biomedical. All other chemicals were reagent grade.

The oligonucleotides used for mutagenesis were purchased from the Penn State Biotechnology Institute. The 15-residue peptide containing the α subunit sequence from Arg89 to Ala103, acetyl-RQKHPTIPIGLLMYA-amide, was purchased from the Protein Structure Facility of the University of Iowa. A 25-residue peptide containing the α subunit sequence from Ala79 to Ala103, acetyl-AQCFEMLALIRQKHPTIPIGLLMYA-amide, was purchased from the Molecular Core Facility of the Pennsylvania State University at the Hershey Medical Center. The purity of the peptides was determined by C18 reversed phase high-pressure liquid chromatography. The purity of the 15-residue peptide exceeded 95%.

*Site-Directed Mutagenesis.* The replacement of Pro93 with Ser (P93S[2] ) in the α subunit was constructed by using the uracil phage method of Kunkel *et al*. (1987). The *Eco*RI/*Sal*I fragment from the vector pXH, a derivative of pTZ18 containing the gene for the α subunit of tryptophan synthase from *E. coli* (Chen & Matthews, 1994), was cloned into the M13mp19 vector using standard molecular biology protocols (Sambrook *et al.*, 1989). The mutant double-stranded DNA was then cloned into the pXH vector for expression of the protein in *E. coli* strain CB149.

The I95A, P96A, I97A, L100A, and Y102A mutants were constructed by the megaprimer polymerase chain reaction (PCR) method (Sarkar & Sommer, 1990) on a modified version of the pBN55 plasmid. The pBN55 plasmid is a derivative of pBR322 containing the gene for the α subunit of tryptophan synthase from *E. coli* (a gift from Dr. Brian Nichols, University of Illinois, Chicago). The resulting plasmids were transformed into *E. coli* strain W3110 (ΔtonB-trp)BA17 *his*$^-$. Expression of the I97A and P96A mutants from the pBN55 plasmid was poor. These mutants were cloned into pXH and transformed into *E. coli* strain CB149 for expression as inclusion bodies.

The H244A mutant and the P93A/H244A double mutant were constructed by the megaprimer method (Sarkar & Sommer, 1990) using the pXH plasmid as a template. For the double mutant P93A/H244A, the plasmid containing the single H244A mutation was used as the template for the second mutation. The resulting plasmids were transformed into *E. coli* strain CB149.

In all cases, the mutations were screened and verified by sequencing the double-stranded DNA for the entire gene using the Sequenase Sequencing kit (U.S. Biochemicals).

*Protein Isolation and Purification.* The I95A, L100A, and Y102A mutants were isolated as soluble protein from *E. coli* strain W3110 (ΔtonB-trp)BA17 *his*$^-$ containing the mutated plasmid pBN55. The P93S, P96A, and I97A mutants and the double mutant P93A/H244A proteins were isolated from inclusion bodies produced in *E. coli* strain CB149 containing the mutated plasmid pXH. The purification procedure for both soluble and insoluble proteins is a modified version of the method of Kirshner *et al*. (1975), previously described by Tsuji *et al.* (1993).

All of the purified mutant proteins were subjected to an activity assay to confirm the presence of a functional α subunit. The purity of each mutant was demonstrated by the observation of a single band on $NaDodSO_4$−polyacrylamide gel electrophoretograms. The proteins were stored in 70% saturated ammonium sulfate, 10 mM potassium phosphate, 4 mM $K_2EDTA$, and 1 mM DTE at pH 7.80 and 4 °C. Under these conditions, no change in the catalytic activity was observed for a period of several months.

*Urea-Unfolding Studies by NMR.* The procedure for the preparation of proteins for proton NMR studies in $^2H_2O$ was the same as described elsewhere for wild type α subunit (Saab-Rincón *et al*., 1993). Three stock protein samples were prepared for each experiment. After the second hydrogen/deuterium exchange and lyophilization step, the samples were dissolved with 1 mL of 56 mM potassium phosphate buffer at pH 7.80 containing 0, 5, or 9 M urea in 100% $^2H_2O$. The pH (uncorrected meter reading) was then adjusted to 7.80 with diluted $^2HCl$ or $NaO^2H$. The final buffer contained 58 mM potassium phosphate, 0.2 mM EDTA, and 2 mM DTE at pH 7.80. After the NMR spectra of these stock samples were collected, appropriate volumes were mixed to provide samples with a constant protein concentration (∼1 mM) and varying urea concentrations. Samples were allowed to equilibrate for 1 h at the final urea concentration before the data were collected.

With one exception, the NMR spectra were obtained at 500 MHz on a Bruker AM 500 spectrometer with the same instrumental settings as previously described for the wild type α subunit (Saab-Rincón *et al*., 1993). The spectra for the P93A/H244A mutant α subunit at 5 °C were collected at 600 MHz on a Bruker DMX 600 spectrometer at the Fox Chase Cancer Research Institute, using the same instrumental settings. The total recycle time between detection pulses was 4-fold greater than measured spin-lattice relaxation times to ensure complete recovery of the signals and accurate peak area measurements.

The number of histidine Cϵ protons associated with each resonance in the region from 7.7 to 8.0 ppm was determined

[2] The mutant protein is designated using the single-letter amino acid code with the first letter corresponding to the wild type residue, the number indicating the position of the residue, and the second letter corresponding to the mutation. P93S indicates the replacement of proline 93 with serine.

from measurements of the peak areas using the program Glinfit (Ontario, Canada). The sum of the areas was normalized to four protons to reflect the four histidines present in wild type α subunit. For those proteins which contain the H244A mutation, the total area was normalized to three protons. The relative area of the resonance due to the His92 Cϵ proton at and above 4 M urea was plotted against urea concentration, and the data were fit to a two-state model (Saab-Rincón *et al*., 1993). Details of the fitting procedure, which assume a linear dependence of the free energy of unfolding on the denaturant concentration (Schellman, 1978), have been previously described (Matthews, 1987).

## RESULTS

Histidine 92 is located in a loop between helix 2 and strand 3 in the α/$\beta$ barrel motif adopted by the α subunit of tryptophan synthase from *Salmonella typhimurium* (Hyde *et al*., 1988). The sequences of the *E. coli* and *S. typhimurium* α subunits differ at 40 of the 268 amino acid positions (Nichols & Yanofsky, 1979). Those differences generally reflect conservative amino acid replacements and do not interfere with the ability of an α subunit from one source to enhance the activity of the $\beta_2$ subunit from the other source (Schneider *et al*., 1981). Pertinent to the present study, the nearest hydrophobic amino acid replacement to His92 occurs at position 52 where valine in *S. typhimurium* is replaced by isoleucine in *E. coli*. Preliminary NMR studies show that the α subunit from *S. typhimurium* undergoes an unfolding transition at high concentrations of urea similar to that of the α subunit from *E. coli* (P. J. Gualfetti and C. R. Matthews, unpublished results). Also, previous results from this lab have shown that α subunits from both strains display the same kinetic patterns during folding (Stackhouse *et al*., 1988). Therefore, it is reasonable to assume that the structures and folding mechanisms of the α subunits from *E. coli* and *S. typhimurium* are very similar.

Inspection of the amino acid sequence near His92 shows a nearly continuous string of nonpolar amino acids, from Ile95 to Tyr102, which define strand 3. Reasoning from the results of Neri *et al*. (1992a) on the residual hydrophobic cluster structure in the 434 repressor, we hypothesized that the residual structure in the α subunit might also involve nonpolar side chains which are constrained to be near His92 by virtue of their proximity in the primary structure. Peptides containing only continuous strings of nonpolar side chains are often difficult to solubilize, making them plausible candidates for residual structure.

This hypothesis was subjected to both thermodynamic and mutagenic tests. Figure 1A shows a series of proton NMR spectra of the histidine Cϵ region of the α subunit at urea concentrations ranging from 4.0 to 9.0 M urea. Global analysis of both the optical and NMR data indicates that I2 and U are the dominant conformers in solution above 4 M urea at 25 °C and pH 7.8 (P. J. Gualfetti, O. Bilsel, and C. R. Matthews, unpublished results). Previous studies have shown that the resonance at 7.8 ppm in 4.0 M urea corresponds to the Cϵ proton of His244 and the minor fraction of the signal from His92 (Saab-Rincón *et al*., 1993) in I2. The small resonance immediately downfield from that for His244 may result from a small amount of irreversibly unfolded protein or from an amide proton which is surprisingly resistant to exchange. Whatever its source, the absence of a dependence of the peak area upon the urea concentration demonstrates that this peak is not sensitive to the I2 ⇄ U transition. The larger, composite resonance at 7.7 ppm corresponds to the signals from His146, His195, and the major fraction of His92. The splitting of the His92 signal into two slowly exchanging components is presumed to reflect the presence of both trans and cis isomers for the peptide bond linking His92 and Pro93 (Saab-Rincón *et al*., 1993). This conjecture is proven to be correct in the present paper (see below).

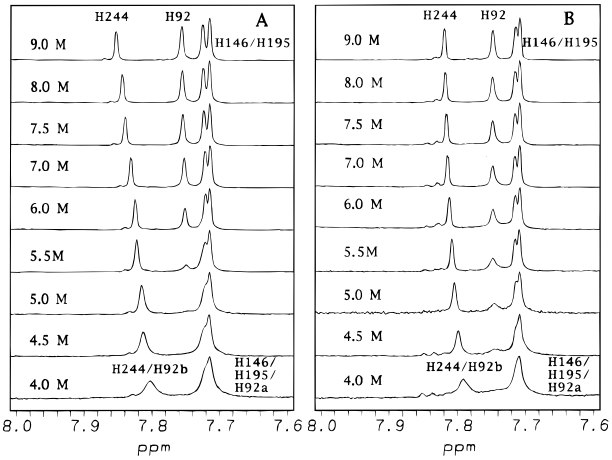


FIGURE 1: FT NMR spectra (500 MHz) of the WT (A) and I97A mutant (B) α subunit at 25 °C and pH 7.8 at increasing concentrations of urea. Spectra were obtained in 58 mM potassium phosphate, 0.2 mM $K_2$EDTA, and 2.0 mM DTE at pH 7.8. The protein concentration was ≈0.8 mM. Chemical shifts are measured relative to DSS. The His146 and His195 resonances have not been assigned to the corresponding residues.

As the urea concentration is increased from 4.0 to 9.0 M, the His244/His92b peak shifts downfield and narrows. The upfield peak resolves into a pair of resonances which, at high urea concentrations, correspond to His146 and His195. Beginning at about 5.0 M urea, a new resonance appears near 7.75 ppm. The area of this resonance, which corresponds to His92, progressively increases as the urea concentration increases to about 7.5 M. Above this urea concentration, the area is comparable to those from the remaining histidine residues. The slow exchange process implied by these peak area changes must have a lifetime greater than ∼40 ms at 6 M urea and 25 °C.

The dependence of the area of the His92 resonance at 7.75 ppm on the urea concentration can be used to estimate the urea dependence of the apparent fraction of unfolded protein, $F_{app}$, and the apparent stability of the intermediate I2 (Aune & Tanford, 1969; Pace, 1986). The results are shown in Figure 2 and Table 1. The sigmoidal change in peak area between 4.0 and 7.5 M urea at 25 °C (Figure 2) is indicative of a cooperative unfolding transition for this intermediate. When these data were fit to a two-state model, it was found that the apparent free energy of unfolding of I2 in the absence of denaturant is 8.2 ± 0.9 kcal mol$^{-1}$, the cooperativity parameter for the transition is 1.4 ± 0.2 kcal mol$^{-1}$ (molar urea)$^{-1}$, and the midpoint is 5.9 ± 0.1 M urea (Saab-Rincón *et al*., 1993). Note that these are apparent thermodynamic parameters because the I2 ⇄ U transition is assumed, but has not been proven, to be a two-state process. Obtaining accurate parameters also requires a global analysis of the N ⇄ I1 ⇄ I2 ⇄ U equilibrium model which describes the

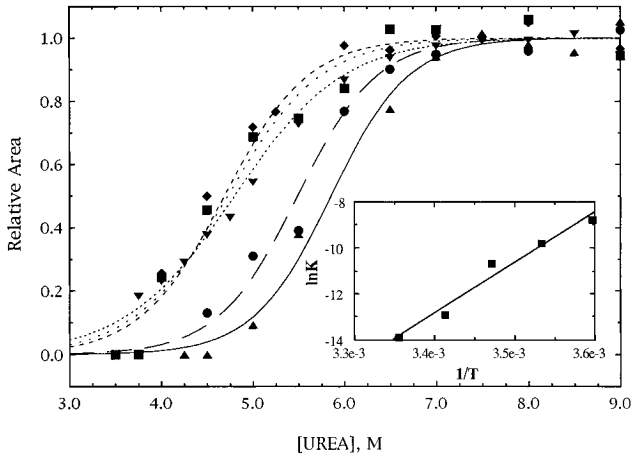


FIGURE 2: Normalized relative area for the His92 Cϵ proton between 3 and 9 M urea for the WT α subunit monitored at 5 (▼), 10 (■), 15 (◆), 20 (●), and 25 (▲) at pH 7.8; estimated errors are ±10%. The inset plot shows the van't Hoff analysis of the data. The solvent is described in Figure 1. The protein concentration was ≈0.8 mM.

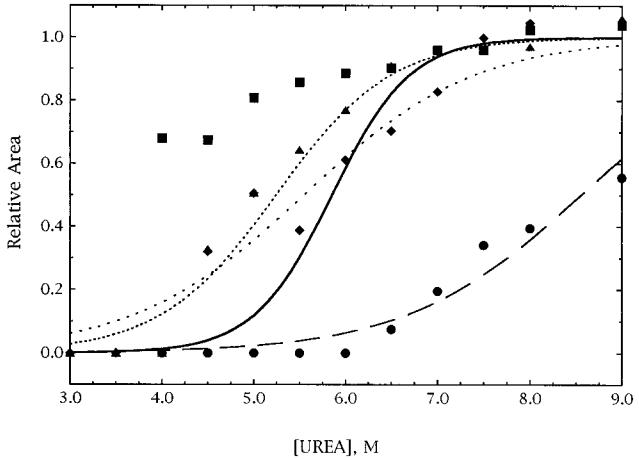


FIGURE 3: Normalized relative areas for the His92 Cϵ protons between 3 and 9 M urea, monitored at 25 °C and pH 7.8 for P93S (◆), I97A (■), P96A (●), and L100A (▲); estimated errors are ±10%. The solvent is described in Figure 1. The protein concentration was ≈0.8 mM. The lines for the P93S, P96A, and L100A data sets represent fits to a two-state model. The solid line represents WT α subunit data collected under the same conditions and fit to a two-state model, as shown in Figure 2.

Table 1: Thermodynamic Parameters for the I2 ⇄ U Transition in Wild Type and Mutant α Subunits

| protein | temp (°C) | $\Delta G^0_{I2/U}$ [kcal/mol] | $-A_{I2/U}$ [kcal/mol$^{-1}$ (molar urea)$^{-1}$] | $Cm_{I2/U}$ [molar urea] |
|---|---|---|---|---|
| WT | 25 | 8.2 (0.9)[a] | 1.4 (0.2) | 5.9 (0.06) |
| | 20 | 7.5 (1.1) | 1.4 (0.3) | 5.4 (0.1) |
| | 15 | 6.1 (1.0) | 1.3 (0.3) | 4.7 (0.1) |
| | 10 | 5.5 (0.9) | 1.2 (0.2) | 4.6 (0.09) |
| | 5 | 4.8 (0.3) | 1.0 (0.07) | 4.8 (0.04) |
| P93S | 25 | 4.1 (0.8) | 0.75 (0.14) | 5.4 (0.2) |
| P93A/H244A | 25 | 3.4 (0.9) | 0.64 (0.15) | 5.4 (0.2) |
| I95A | 25 | 7.9 (1.5) | 1.29 (0.25) | 6.2 (0.1) |
| P96A | 25 | 4.9 (0.5) | 0.56 (0.07) | 8.6 (0.1) |
| I97A | 25 | nd[b] | nd | <4.0 |
| L100A | 25 | 4.7 (0.6) | 0.90 (0.12) | 5.2 (0.2) |
| Y102A | 25 | 5.4 (0.6) | 1.06 (0.12) | 5.0 (0.1) |
| H244A | 5 | 4.0 (1.7) | 0.82 (0.34) | 4.8 (0.2) |

[a] Errors are represented in parentheses and correspond to one standard deviation. [b] Not determined.

complete thermodynamic behavior of the α subunit in the presence of increasing concentrations of urea. This analysis is in progress (P. J. Gualfetti, O. Bilsel, and C. R. Matthews, unpublished results).

The possibility that the I2 intermediate is an associated species can be ruled out by inspection of the NMR spectra in the range from 5 to 9 M urea (Figure 1A). The line widths of the histidine resonances do not change as I2 is converted to the monomeric, unfolded form, consistent with a monomeric form for I2. Note that the better resolution of the two upfield resonances at increasing urea concentrations reflects a different urea dependence of their chemical shifts (Saab-Rincón *et al*., 1993).

The effect of temperature on this urea-induced unfolding reaction is shown in Figure 2 and Table 1. The apparent stability of I2 monotonically decreases as the temperature is lowered from 25 to 5 °C. This decrease in the stability with decreasing temperature is consistent with an exothermic reaction, i.e., $\Delta H < 0$. The slope of the His92 transition also decreases as the temperature is lowered, resulting in a nonmonotonic change in the midpoint at 5 °C. A van't Hoff analysis of the apparent free energies of unfolding calculated from these data (Figure 2, inset) yielded an apparent enthalpy change in the absence of urea of $-44 \pm 4$ kcal mol$^{-1}$. When considered with the apparent free energy change, +8.2 kcal mol$^{-1}$ (Table 1), these data show that the apparent entropy change for the unfolding of I2 is $-175 \pm 17$ cal mol$^{-1}$ K$^{-1}$ at 25 °C. Negative values for the enthalpy and entropy changes imply that I2 is stabilized by the hydrophobic effect (Kauzmann, 1959).

Given this evidence for a primary role for the hydrophobic effect in stabilizing I2, a series of mutations were constructed in the amino acid sequence from Pro93 to Tyr102. This region corresponds closely to the third $\beta$ strand in the α subunit, a segment which is almost fully buried in the native conformation (Hyde *et al*., 1988). A series of NMR spectra of the histidine region for the I97A mutant α subunit are shown in Figure 1B. The dependence of these spectra on the urea concentration is qualitatively similar to that for the wild type protein. However, direct comparison of the spectra collected at 5.0, 5.5, and 6.0 M urea shows that the area of the His92 peak for the mutant protein is substantially larger at the same urea concentration than that for the wild type protein. This result implies that the I2 intermediate for the I97A α subunit is destabilized by the mutation. The urea dependence of the apparent fraction of unfolded I97A α subunit, determined from the area of the His92 resonance, is shown in Figure 3. The $F_{app}$ curve could only be determined with confidence above 4.0 M urea because the line widths of all of the histidine resonances are broadened by the self-association of another stable intermediate, I1, below this urea concentration (Figure 1 and J. C. Lee, G. Saab-Rincón, and C. R. Matthews, unpublished results). This limitation prohibited an accurate fit of the I97A data to a two-state model. Even in the absence of a fit, however, it is evident that the midpoint of the I2 ⇄ U transition for the I97A protein is less than 4.0 M urea.

NMR spectra were also collected for the P93S, P93A/H244A (double mutant), I95A, P96A, L100A, Y102A, and H244A mutant α subunits at urea concentrations between 4.0 and 9.0 M. The single and double mutants containing H244A were constructed to test whether the minor resonance from the His92 proton is, as was previously assumed (Saab-Rincón *et al*., 1993), coincident with the resonance from

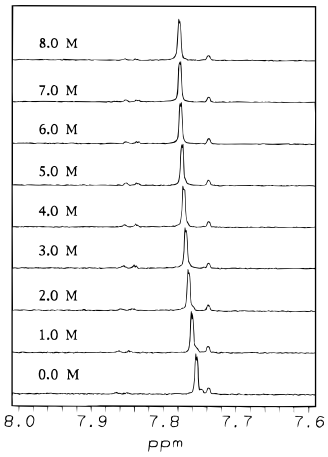


FIGURE 4: FT NMR spectra (500 MHz) of the peptide acetyl-RQKHPTIPIGLLMYA-amide at 25 °C and pH 7.8 at increasing concentrations of urea. The solvent is described in Figure 1. The peptide concentration was ≈1.0 mM.

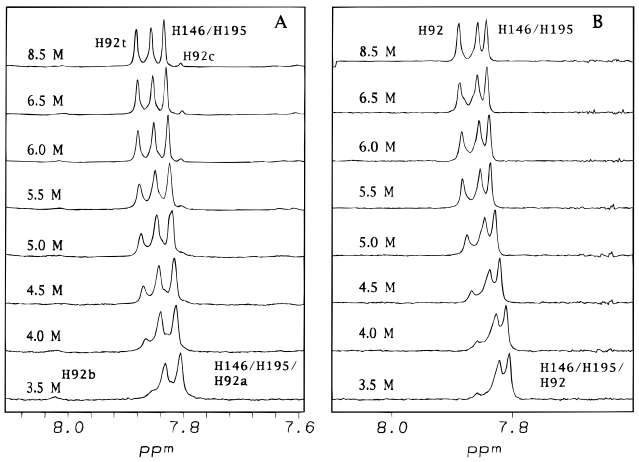


FIGURE 5: (A) FT NMR spectra (500 MHz) of the H244A mutant and (B) FT NMR spectra (600 MHz) of the P93A/H244A double mutant of the α subunit at 5 °C and pH 7.8 at increasing concentrations of urea. The solvent is described in Figure 1. The protein concentration was ≈0.8 mM.

His244. The urea dependence of the area of the His92 resonance was reasonably well described by a two-state model in all cases; $F_{app}$ curves for P93S, P96A, and L100A are shown in Figure 3. The thermodynamic parameters extracted from all of the fits are given in Table 1.

With the exception of the I95A and H244A mutations, all of these replacements perturb the unfolding of the I2 species. Note that the H244A data were collected at 5 °C to increase the dispersion of the histidine resonances. L100A and Y102A have lower midpoints, cooperativity parameters, and stabilities than wild type protein. P93S, P93A/H244A, and P96A all have lower stabilities and substantially lower cooperativity parameters. The midpoint for P96A is significantly higher than that for wild type α subunit, while the midpoints for the two Pro93 replacements are lower. Although the effects differ from mutant to mutant, it is evident that all of these side chains, with the possible exception of Ile95 and His244, are involved in the stabilization of the residual structure which defines the I2 intermediate in the wild type α subunit. Other mutations at positions 95 and 244 are required to prove definitively that these side chains do not participate in the residual structure.

To test whether the side chains at positions 93, 96, 97, 100, and 102 are sufficient to determine the structure of I2, a 15-residue peptide encompassing the region from Arg89 to Ala103 was examined. NMR spectra of this peptide as a function of urea concentration over the range from 0 to 8.0 M urea are shown in Figure 4. The major peak at 7.75 ppm displays a slight downfield shift as the urea concentration increases from 0 to 5.0 M urea. The minor peak at 7.74 ppm, which is probably due to the cis isomer of the His92−Pro93 peptide bond (see below), is insensitive to the urea concentration. The absence of the slow exchange process observed in full length protein above 4.0 M urea demonstrates that the residual structure in the full length protein must involve side chains other than those in strand $\beta_3$ that were mutated.

To examine the possible role of nonpolar residues in helix 2 in I2, a 25-residue peptide encompassing the region from Ala79 to Ala103 was also synthesized. This sequence contains the amphipathic sequence in helix 2, the intervening turn including His92, and the hydrophobic strand $\beta_3$. Unfortunately, this peptide aggregated in water at concentrations required for NMR spectroscopy, precluding an analysis of the role of $\alpha_2$ in the residual structure.

The role of Pro93 in the perturbation of the signal from the C$\epsilon$ proton of His92 in both intermediate and unfolded forms of the α subunit was investigated by comparison of the NMR spectra of the H244A and P93A/H244A mutant proteins at a series of urea concentrations (Figure 5). The replacement of His244 with Ala makes it possible to observe whether a minor, downfield resonance is actually associated with His92 in the I2 intermediate; these two resonances were assumed to be coincident in wild type α subunit (Saab-Rincón *et al.*, 1993). These spectra were collected at 5 °C to better resolve the remaining histidine resonances and to minimize the self-association of the I1 intermediate. Evident in the H244A spectrum (Figure 5A) at 3.5 M urea is a small, broad resonance at 8.02 ppm (labeled H92b) whose area decreases as the urea concentration increases. This peak diminishes in concert with the increase in area of a small peak at 7.8 ppm (labeled H92c). This latter peak represents ∼5% of the amplitude of the individual major peaks at high urea concentrations. The relative magnitude and upfield shift are similar to those for the minor peak observed in the 89−103 peptide (Figure 4), suggesting that it reflects the cis isomer of the His92−Pro93 peptide bond.

The assignment of both of these resonances to the minor isomer (presumably the cis form) of the His92−Pro93 peptide bond was made by examination of the urea titration of the P93A/H244A double mutant under the same conditions (Figure 5B). Both minor peaks are absent at all urea concentrations from 3.5 to 8.5 M urea, demonstrating that the existence of these alternative intermediate and unfolded forms depends upon the presence of the proline residue at position 93.

## DISCUSSION

The results of thermodynamic and mutational experiments on the residual structure contained in the I2 intermediate of the α subunit of tryptophan synthase from *E. coli* are consistent with the association of nonpolar side chains in a hydrophobic cluster which is stable in 5 M urea. This cluster exists in the absence of well-defined secondary structure, as measured by the far-UV CD spectrum, and tertiary structure,

as measured by the absorbance spectrum (Saab-Rincón *et al*., 1993). The observation that a 15-residue peptide, which includes the five positions where replacements perturbed the stability of I2, was not sufficient to define the residual structure implies that other, possibly nonlocal interactions are involved. The insolubility of a 25-residue peptide that also included the preceding $\alpha_2$ helix precluded an examination of its potential role in the stabilization of I2.

The absence of secondary structure in I2 clearly demonstrates that this intermediate is not a molten globule (Kuwajima, 1989) or a more extended version of a molten globule that has been reported for $\beta$-lactamase at 4 °C (Uversky & Ptitsyn, 1994). The question arises for the I2 intermediate in the α subunit of how a folded form of a protein can be stabilized by the hydrophobic effect in the apparent absence of secondary structure (as measured by the far-UV CD spectrum). The answer may be found in a study by Koide *et al*. (1993), who examined the folding of apoplastocyanin with a variety of techniques. These investigators observed a transient intermediate in which 16 $\beta$ sheet amide protons are protected from exchange with solvent. Although this species has one or more of its tyrosines excluded from solvent, it does not bind ANS and it has no appreciable signal in the far-UV CD spectrum. Consistent with the CD results is a 1D NMR spectrum which shows few resonances in the region expected for the Cα protons of residues in $\beta$ structures (5−6 ppm). These results were interpreted to mean that at least a part of the $\beta$ sheet hydrogen-bonding pattern found in native apoplastocyanin is also present in this intermediate. However, in contrast to the native form, the sheet is sufficiently flexible and dynamic to reduce the amplitude of the far-UV CD spectrum and shift the Cα resonances toward those frequencies expected for a random coil.

Assuming that this interpretation of the apoplastocyanin results can be extended to the I2 intermediate in the α subunit, the lack of a signal in the far-UV CD spectrum of the α subunit at 5 M urea may only indicate the absence of extensive, well-defined secondary structure. Localized, fluctuating backbone−backbone hydrogen bonding might still occur, making it possible to form an extensive hydrophobic cluster(s) between the pendant side chains. The putative dynamic nature of the backbone in this type of hydrophobic cluster differentiates it from that found in a molten globule and may explain why the intermediate in apoplastocyanin does not bind ANS. Perhaps only a small fraction of the ensemble of structures accessible to this species has the type of stable hydrophobic surface which favors ANS binding and which is prominent in a molten globule. This stable surface presumably reflects the underlying, well-defined secondary structure which is characteristic of molten globules. Preliminary results indicate that the I2 intermediate in the α subunit, similar to the apoplastocyanin intermediate, does not bind ANS (J. A. Zitzewitz and C. R. Matthews, unpublished results).

The negative sign for the enthalpy difference between the I2 and U states demonstrates, along with the negative entropy change, that this species is stabilized principally by the hydrophobic effect (Kauzmann, 1959). The magnitude of the enthalpy and entropy changes for the α subunit is surprisingly large, comparable to the values measured by calorimetry for the thermal unfolding of many proteins (Privalov & Gill, 1989). A quantitative interpretation of these parameters and the effects of mutations on the apparent stability of the I2 intermediate, however, is not warranted until the two-state model assumed to describe the I2 ⇄ U transition is verified.

Assuming for the present, however, that these parameters are at least semiquantitatively correct, the large decreases in both enthalpy and entropy imply that a substantial fraction of the nonpolar side chains in I2 associate with each other in a polypeptide whose backbone and side chains are very mobile. This intermediate behaves as if 5 M urea has disrupted a highly organized, long range hydrogen-bonding network without simultaneously solubilizing the nonpolar side chains. The negative value for the entropy change implies that the gain in backbone and side chain entropy upon the unfolding of the I2 intermediate is more than offset by the loss in the entropy of the solvent as it organizes about the nonpolar side chains in the unfolded form.

The I2 intermediate in the α subunit is sufficiently stable in the absence of urea that this species could play a significant role in the actual folding mechanism. However, the fact that I2 is in slow exchange with the unfolded form at 6 M urea suggests that it may not be kinetically accessible in the same time range for which secondary structure is known to appear. Stopped-flow far-UV CD studies on the α subunit have shown that an intermediate with significant ellipticity appears within 5 ms following the initiation of the refolding reaction (Ogasahara & Yutani, 1994; G. Saab-Rincón, P. J. Gualfetti, and C. R. Matthews, unpublished results). The slow exchange behavior for the His92 Cϵ proton is consistent with a reaction whose relaxation time near 6 M urea exceeds 4 ms by at least 1 order of magnitude (Saab-Rincón *et al*., 1993). A 40 ms or longer relaxation time for the I2 ⇄ U transition at 6 M urea, however, does not rule out a substantially faster refolding reaction under strongly folding conditions. The maximum value for a protein-folding relaxation time occurs at the midpoint of the transition; the relaxation time decreases exponentially as the denaturant concentration either increases or decreases (Matthews, 1987).

Supporting this contention are the results of stopped-flow far-UV CD refolding experiments which begin at 5, 6, and 7 M urea and end at 0.5 M urea (G. Saab-Rincón, P. J. Gualfetti, and C. R. Matthews, unpublished results). All three traces show the same burst phase CD signal. Because the relative concentration of the I2 species decreases from 65 to less than 12% from 5 to 7 M urea (P. J. Gualfetti and C. R. Matthews, unpublished results), these results are consistent with the formation and further folding of the I2 intermediate to the CD detectable I1 intermediate within 5 ms at 0.5 M urea. Therefore, the thermodynamic stability and kinetic accessibility of I2 make it a potential intermediate on the kinetic pathway of folding of the α subunit. Agashe *et al.* (1995) have recently reported that barstar initially collapses to a compact state with no secondary or specific tertiary structure, consistent with this hypothesis for the role of I2 in the folding of the α subunit.

The question arises as to why residual structure persists in the α subunit of tryptophan synthase, the 434 repressor (Neri *et al*., 1992b), staphylococcal nuclease (Shortle & Meeker, 1989), and the fatty acid binding protein (Ropson & Frieden, 1992) when these proteins are dissolved in high concentrations of denaturant. An intriguing possibility is that continuous strings of nonpolar side chains, such as those

found in $\beta$ strands, tend to associate in hydrophobic clusters rather than behave as random chains in denaturant/water solutions. Although the 434 repressor is a helical protein, the side chains of Val54, Val56, Trp58, and Leu59 which define its residual structure are in close proximity in the sequence. Each of the three remaining examples contains five or more $\beta$ strands which could nucleate the formation of hydrophobic clusters. Matheson and Scheraga (1978) have previously suggested that stretches of nonpolar side chains could act as chain-folding initiation sites.

If this hypothesis is correct, residual structure may exist in many other proteins which are dissolved in high concentrations of chemical denaturant. This structure might not be apparent in any of the standard optical spectra but could be detected by high-resolution NMR spectroscopy (Neri *et al*., 1992a; Ropson & Frieden, 1992). Extension of these NMR methods to unfolded proteins (Logan *et al*., 1993) should improve our understanding of their structures and, thereby, enhance the analysis of thermodynamic and kinetic folding mechanisms.

## ACKNOWLEDGMENT

We thank Dr. Heinrich Roder for providing access to the 600 MHz NMR at Fox Chase Cancer Research Institute and Dr. Yuzhu Zhong for his technical assistance. We also thank Ieva Perkons and Matt Houser for their assistance with the manuscript.

## REFERENCES

Agashe, V. R., Shastry, M. C. R., & Udgaonkar, J. B. (1995) *Nature 377*, 754−757.
Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M., & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9412−9416.
Aune, K. C., & Tanford, C. (1969) *Biochemistry 8*, 4579−4585.
Buck, M., Radford, S. E., & Dobson, C. M. (1994) *J. Mol. Biol. 237*, 247−254.
Chen, X., & Matthews, C. R. (1994) *Biochemistry 33*, 6356−6362.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
Flory, P. J., & Miller, W. G. (1966) *J. Mol. Biol. 15*, 284−297.
Griko, Y. V., Freire, E., & Privalov, P. L. (1994a) *Biochemistry 33*, 1889−1899.
Griko, Y. V., Gittis, A., Lattman, E. E., & Privalov, P. L. (1994b) *J. Mol. Biol. 243*, 93−99.
Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., & Davies, D. R. (1988) *J. Biol. Chem. 263*, 17857−17871.
Jacobs, M. D., & Fox, R. O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 449−453.
James, E., Wu, P. G., Stites, W., & Brand, L. (1992) *Biochemistry 31*, 10217−10225.
Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1−63.
Kirschner, K., Wiskocil, R. L., Foehn, M., & Rezeau, L. (1975) *Eur. J. Biochem. 60*, 513−523.
Koide, S., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 12299−12310.
Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87−103.
Logan, T. M., Olejniczak, E. T., Xu, R. X., & Fesik, S. W. (1993) *J. Biomol. NMR 3*, 225−231.
Matheson, R. R. J., & Scheraga, H. A. (1978) *Macromolecules 11*, 819−829.
Matthews, C. R. (1987) *Methods Enzymol. 154*, 498−511.
Matthews, C. R. (1993) *Annu. Rev. Biochem. 62*, 653−683.
Neri, D., Billeter, M., Wider, G., & Wüthrich, K. (1992a) *Science 257*, 1559−1563.
Neri, D., Wider, G., & Wüthrich, K. (1992b) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4397−4401.
Nichols, B. P., & Yanofsky C. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5244−5248.
Ogasahara, K., & Yutani, K. (1994) *J. Mol. Biol. 236*, 1227−1240.
Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
Privalov, P. L., & Gill, S. J. (1989) *Adv. Protein Chem. 39*, 191−234.
Reynolds, J. A., & Tanford, C. (1970) *J. Biol. Chem. 245*, 5161−5165.
Ropson, I. J., & Frieden, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7222−7226.
Saab-Rincón, G., Froebe, C. L., & Matthews, C. R. (1993) *Biochemistry 32*, 13981−13990.
Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sarkar, G., & Sommer, S. S. (1990) *BioTechniques 8*, 404−407.
Schellman, J. A. (1978) *Biopolymers 17*, 1305−1322.
Schneider, W. P., Nichols, B. P., & Yanofsky C. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2169−2173.
Shalongo, W., Jagannadham, M., & Stellwagen, E. (1993) *Biopolymers 33*, 135−145.
Shortle, D., & Meeker, A. K. (1989) *Biochemistry 28*, 936−944.
Stackhouse, T. M., Onuffer, J. J., Matthews, C. R., Syed, A. A., & Miles, E. W. (1988) *Biochemistry 27*, 824−832.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121−282.
Tsuji, T., Chrunyk, B. A., Chen, X., & Matthews, C. R. (1993) *Biochemistry 32*, 5566−5575.
Uversky, V. N., & Ptitsyn, O. B. (1994) *Biochemistry 33*, 2782−2791.

BI951726O