Real-Time Refolding Studies of 6-^{19}F-Tryptophan Labeled *Escherichia coli* Dihydrofolate Reductase Using Stopped-Flow NMR Spectroscopy†

Sydney D. Hoeltzli and Carl Frieden*

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received July 31, 1996; Revised Manuscript Received October 8, 1996⊗

ABSTRACT: *Escherichia coli* dihydrofolate reductase (ecDHFR, EC1.5.1.3) contains 5 tryptophan residues that have been replaced with 6-^{19}F-tryptophan. Five native and four of the five unfolded tryptophan resonances can be resolved in the 1D ^{19}F NMR spectra and have been assigned [Hoeltzli, S. D., & Frieden, C. (1994) *Biochemistry 33*, 5502−5509]. This resolution allows the behavior of the native and the unfolded resonances assigned to each individual tryptophan to be monitored during the unfolding or refolding process. We now use these assignments and stopped-flow NMR to investigate the real-time behavior of specific regions of the protein during refolding of DHFR after dilution from 4.6 to 2.3 M urea (midpoint of the transition = 3.8 M) at 5 °C. Approximately half of the intensity of each of the four unfolded resonances is present at the first measurable time point (1.5 s). Little native resonance intensity is detectable at this time. The remaining unfolded resonance intensities present then disappear in two phases, with rates similar to the two slowest phases observed by either stopped-flow fluorescence or circular dichroism spectroscopy upon refolding under the same conditions. Substantial total resonance intensity is missing during the first 20 s of the refolding process. The appearance of the majority of native resonance intensity (as assessed by the height of each of the five native tryptophan resonances) is slow and similar for all five tryptophans. In contrast, the largest amplitude changes observed by either stopped-flow far-UV circular dichroism spectroscopy or fluorescence spectroscopy, and the greatest loss of unfolded resonance intensity, occur much more rapidly. We conclude from these studies: (1) that, under these conditions, the unfolded state remains substantially populated after initiation of refolding; (2) that the early steps in refolding involve a solvent protected intermediate containing substantial secondary structure, but (3) that the stable native side chain interactions form slowly and are associated with the final rate-limiting phase of the refolding process. Preliminary analysis of the area of broadened native resonances suggests that these resonances may appear at different rates, indicating that some regions of the protein begin to sample a native-like side chain environment while side chain environment in other regions of the protein remains less ordered. The results of this study are consistent with the earlier studies demonstrating that mobility of side chains is an early step in unfolding [Hoeltzli, S. D., & Frieden, C. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9318−9322] and that recovery of enzymatic activity occurs as a late step in the folding process [Frieden, C. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4413−4416].

The mechanism by which a protein folds to its correct tertiary structure remains one of the key unanswered questions in biochemistry. Common experimental approaches measure rates of unfolding or refolding by monitoring changes in intrinsic fluorescence, absorbance, or circular dichroism. When multiple phases are observed, they may represent either separate pathways or the formation of intermediates on a single pathway. Fluorescence, absorbance, or circular dichroism measurements do not readily distinguish between these mechanisms. If intermediates exist, these techniques cannot provide specific information about the structure of such intermediates.

A number of studies have identified specific regions of secondary structure formed during the folding process by using hydrogen−deuterium exchange combined with multidimensional proton NMR spectroscopy (e.g., Englander & Mayne, 1992; Baldwin, 1993; Roder, 1989). This technique, by monitoring rates of amide proton exchange, provides detailed and specific information about the behavior of the backbone of the protein and the formation of secondary structure but does not, in general, monitor side chain environment. Thus, structural information about the formation of specific regions of tertiary structure and about behavior of side chains during the folding process is still limited.

As a complementary alternative technique, we have been studying protein unfolding (Hoeltzli & Frieden, 1995) and folding by ^{19}F NMR spectroscopy in real time, following changes in side chain environment rather than the formation of secondary structure. A similar rapid mixing ^{1}H NMR experiment has been used to study the unfolding of ribonuclease A (Kiefhaber & Baldwin, 1995; Kiefhaber et al., 1995) and the folding of bovine α-lactalbumin in the absence of calcium (Balbach et al., 1995), by injecting the protein into the NMR tube through narrow tubing. However, these investigators could follow only a few isolated resonances in the 1D ^{1}H spectrum of the native protein. In addition, a number of transients at different times after mixing were averaged to obtain sufficient signal to noise, limiting the

† This work was supported by National Institutes of Health Grant DK13332.

* To whom correspondence should be addressed.

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

S0006-2960(96)01896-X CCC: $12.00

technique to proteins which fold extremely slowly. Because our stopped-flow NMR cell allows different injections to be repeated and same time interval averaged, we can follow a somewhat faster process. In addition, because fluorine chemical shifts are extremely sensitive to environment, we can resolve and follow individual resonances in the unfolded as well as in the native state.

We have applied this stopped-flow NMR method to *Escherichia coli* dihydrofolate reductase (DHFR).[1] This enzyme is a monomer of 159 amino acids and molecular weight 17 680. It catalyzes the NADPH dependent reduction of 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate. Its small size, well-characterized enzyme mechanism (Fierke et al., 1987; Penner & Frieden, 1987), and well-refined structure (Bolin et al., 1982; Bystroff et al., 1990; Bystroff & Kraut, 1991) as well as the reversibility of its folding reaction in the presence of chemical denaturants (Frieden, 1990; Touchette et al. 1986) make it a good model for protein folding studies.

E. coli DHFR contains five tryptophan residues located in different elements of secondary structure and spatially distributed throughout the protein. We have previously prepared 6-^{19}F-tryptophan labeled *E. coli* DHFR, assigned the resonances observed in the ^{19}F spectrum of this protein to individual tryptophans, and studied its behavior at equilibrium in the presence of chemical denaturant (Hoeltzli & Frieden, 1994). Unlabeled and 6-^{19}F-tryptophan labeled DHFR have similar CD spectra and enzymatic properties, indicating that the incorporation of the fluorine label into *E. coli* DHFR is not a structurally perturbing change (Hoeltzli & Frieden, 1994). We have also monitored the changes in the ^{19}F spectrum in real time during denaturant-induced unfolding (Hoeltzli & Frieden, 1995) using a stopped-flow device incorporated into the NMR spectrometer (Hoeltzli et al., 1994). From those data, we concluded that the major phase of unfolding involves the formation of an intermediate that retains substantial secondary structure, but in which the side chains possess considerable mobility. Following the formation of this intermediate, all regions of the protein then unfold at equal rates.

In this paper, we apply stopped-flow NMR spectroscopy to study the behavior of side chains during the urea-induced folding process by monitoring the real-time changes in the NMR spectrum of 6-^{19}F-tryptophan labeled *E. coli* DHFR. We show that the majority of DHFR refolds via at least one intermediate which contains native-like secondary structure, but in which side chains retain considerable mobility until late in the folding process, when the structure becomes stabilized.

MATERIALS AND METHODS

Materials. Ultrapure urea and isopropyl β-D-thiogalactopyranoside were purchased from United States Biochemical (Cleveland, OH). Urea stocks were either prepared the day of use or stored in aliquots at −70 °C until the day of use. The concentration of urea was determined by refractive index at 25 °C using quantitative relationships between concentration and refractive index (Pace, 1986). 6-^{19}F-tryptophan, methotrexate, folate, and methotrexate agarose were obtained from Sigma (St. Louis, MO). Fast flow DEAE Sepharose was from Pharmacia (Piscataway, NJ). All other chemicals were reagent grade.

Protein Labeling and Purification. DHFR containing 6-^{19}F-tryptophan was purified from 5 L fermentations of the *E. coli* auxotroph W3110TrpA33 containing the plasmid pTrc99DHFR. This plasmid was constructed by inserting the *folA* gene from plasmid pTY1 into the plasmid pTrc99A (Pharmacia Co., Piscataway, NJ). The cells were grown in a Biostat B fermentor (Braun Instruments, Allentown, PA) at 37 °C on M9 minimal medium prepared with twice the normal concentration of phosphate salts and supplemented with 1.5 g/L CSM-TRP (Bio-101 Inc., Vista, CA), 0.2% glucose, 0.2 mM L-tryptophan, 50 μg/L ampicillin, and 1 mL/L Poly-Vi-Sol vitamin drops with iron (Mead-Johnson, Evansville, IN) and maintained at pH 7 by addition of NH_4-OH. Glucose was maintained between 0.1% and 0.2% and pO_2 between 25% and 35% by varying the rate of addition of a feed mixture containing 45 g/L CSM-TRP, 1.2 g/L L-tryptophan, 30 g/L NH_4Cl, 4.8g/L $MgSO_4$, and 20% glucose. The cells were harvested at OD_{600} = 16 and resuspended in fresh medium containing 6-^{19}F-tryptophan in place of L-tryptophan. After 30 min of growth in the 6-^{19}F-tryptophan medium, the plasmid was induced with 1 mM isopropyl β-D-thiogalactopyranoside for 2 h and then harvested.

The protein purified from these fermentations was >90% labeled with 6-^{19}F-tryptophan as measured by comparison of deconvoluted resonance intensity to the intensity of a known reference. The protein was purified and a substoichiometric amount of tightly bound folate removed by denaturation and renaturation as previously described (Hoeltzli & Frieden, 1994). The protein was stored in 80% ammonium sulfate until the day prior to use and then dialyzed or diafiltered into the buffer used in the experiment. The enzyme concentration was determined by active site titration of intrinsic fluorescence using methotrexate, which forms a 1:1 complex, and was within 5% of that determined by absorbance (Baccanari et al., 1975) after the removal of residual folate.

For fluorescence and circular dichroism experiments, the protein was dialyzed against a buffer containing 50 mM potassium phosphate, 100 mM KCl, 0.1 mM EDTA, and 1 mM DTT and diluted to the desired enzyme concentration (10−40 μM). For NMR experiments, the protein was exchanged into buffer by diafiltration, concentrated through a YM10 membrane (Amicon, Beverly, MA), and diluted with D_2O and 20 mM 4-^{19}F-phenylalanine to achieve a final buffer concentration of 50 mM potassium phosphate, pH 7.2, 100 mM KCl, 0.1 mM EDTA, 15 mM DTT, 15% D_2O, and 0.3 mM 4-^{19}F-phenylalanine and a final protein concentration of 2.6 mM. No correction to pH was made for D_2O content.

Stopped-Flow Fluorescence and Circular Dichroism Spectroscopy. Stopped-flow fluorescence studies were performed at 5 °C using an Applied Photophysics (Surrey, U.K.) spectrophotometer in the fluorescence mode with a path length of 0.2 cm. Equal volume drive syringes of 2.5 mL were used. Five hundred data points were collected in 5 s, and 500 additional points over the next 500 s. An excitation wavelength of 290 nm was used, and fluorescence was observed using a 305 nm cutoff filter. Five individual injections were averaged and analyzed. Averages of different injections collected on different days produced comparable rates.

[1] Abbreviations: DHFR, *Escherichia coli* dihydrofolate reductase, EC 1.5.1.3; NADPH, reduced nicotinamide adenine dinucleotide phosphate; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid.

The time course of circular dichroism changes was measured at 5 °C using an Applied Photophysics RX1000 rapid kinetics accessory equipped with a 0.2 cm path length stopped-flow cell, equal volume drive syringes of 2.5 mL, and fitted to a JASCO J600 circular dichroism spectrophotometer. Twelve hundred data points were collected at either 20 ms or 0.2 s intervals and combined. Five individual injections were averaged and analyzed. Averages from injections collected on different days produced comparable rates.

Stopped-Flow NMR Spectroscopy. An Applied Photophysics RX1000 rapid kinetics spectrometer accessory was adapted for use as previously described (Hoeltzli & Frieden, 1995; Hoeltzli et al., 1994). Details of the design of the cell and the mixer are available upon request. At 5 °C, 0.8 mL displaces approximately 90% of the old solution from the coil of the probe. The dead time of the apparatus was estimated at 100 ms.

NMR data were collected at 5 °C on a Varian VXR-500 spectrometer operating at 470.3 MHz for ^{19}F using a Nalorac proton/fluorine probe. All spectra were referenced to an internal standard of 4-^{19}F-phenylalanine. Chemical shifts are given relative to external trifluoroacetic acid. Data were acquired using a modification of the Varian VNMR s2pul pulse sequence (Hoeltzli et al., 1994) that delays acquisition for a user defined interval following an external trigger synchronized to sample injection and mixing. After injection and rapid mixing, and an initial delay of 1.5 s, 20 transients were collected 1.03 s apart; 20 transients were collected 3.2 s apart; and 60 transients were collected 6.9 s apart for a total time of 494 s. Data for each transient were acquired for 0.503 s using a 90° pulse and a spectral width of 5999 Hz. Sixty-five separate injections were made. Periodically, an equilibrium spectrum of refolded protein was acquired following an additional 5 min delay (800 s after initiating refolding). Recycle times of less than 4 times the longest spin−lattice relaxation time (T_1), as previously determined (Hoeltzli & Frieden, 1994), were corrected for T_1 assuming that $T_1 \gg T_2$.

The resulting arrayed free induction decays were summed using a C-shell macro which increments through the number of injections for each time interval, making use of the VNMR addfid macro as previously described (Hoeltzli & Frieden, 1995).

Protein samples at the concentration used in the NMR experiment were examined for aggregation by sedimentation velocity studies using an Optima XL-A analytical ultracentrifuge (Beckman, Palo Alto, CA) and by light scattering monitored at 540 nm using a fluorometer (Photon Technology Inc., So. Brunswick, NJ). Samples of concentrated protein denatured at 1.3 mM and incubated in urea for less than 2 h in the presence of a 5-fold excess of DTT showed no evidence of aggregation. During the course of the stopped-flow NMR experiment, protein was denatured for less than 2 h. Active site titration of the refolded material used in the NMR experiment indicated that >85% of the methotrexate binding capacity was recovered.

Data Analysis. Exponential fits were obtained using the program Kaleidagraph (Synergy Software, Reading, PA).

RESULTS

Stopped-Flow NMR Spectroscopy. We have previously assigned the five resonances observed in the ^{19}F NMR spectrum of the native state and the four resonances observed in the ^{19}F NMR spectrum of the unfolded state to the five individual tryptophan residues in *E. coli* DHFR (Hoeltzli & Frieden, 1994). The native chemical shifts are affected by a decrease in temperature from 20 to 5 °C, with the most significant change occurring to Trp30 (−0.34 ppm) and Trp74 (+0.36 ppm) (data not shown). These chemical shift changes were confirmed by examining the spectrum of the previously created Trp74Phe and Trp30Phe mutants. Because five of five native and four of five unfolded tryptophan resonances have distinct chemical shifts, we can monitor changes in both native and unfolded environment in real time during refolding.

Table 1: Refolding Monitored by Fluorescence, Far-UV CD (222 nm), and Disappearance of Unfolded Resonances

	A_1	k_1	A_2	k_2	A_3	k_3
fluorescence[a]	0.102	0.682	0.651	0.151	0.228	0.007
circular dichroism[a]		nd	0.812	0.115	0.204	0.005
^{19}F NMR[b]						
Trp22	0.50[c]	nd[c]	0.30	0.12	0.20	0.007
Trp30 and -47	0.47[c]	nd[c]	0.29	0.16	0.24	0.007
Trp74	0.50[c]	nd[c]	nd[d]	nd[d]	0.32	0.006
Trp133	0.46[c]	nd[c]	0.37	0.26	0.17	0.010

[a] Amplitudes for fluorescence and CD data are given as the ratio of the amplitude of each individual phase to the total amplitude. Data were fit to a double or triple exponential equation using the program Kaleidagraph (Synergy Software, Reading, PA). Estimated error less than 5%. [b] Estimated experimental and fitting error of about 20%. [c] The amplitude of the first phase (A_1) was determined as $1 - (A_2 + A_3)$. Because of the length of the initial delay, this value includes approximately 25% of the amplitude of the second phase. The rate was too rapid to be determined by this technique. [d] nd not determined. Because Trp74 is the least resolved of the four unfolded resonances, the rate constant associated with this phase could not be determined due to uncertainty in the initial time points.

The conditions used in the stopped-flow NMR experiment (5 °C, refolding from 4.6 to 2.3 M urea) were chosen to slow refolding as much as possible. The midpoint for denaturation by urea under the buffer conditions used in this study is unaffected by temperature between 20 and 5 °C. At higher temperatures and lower final concentrations of urea, four phases are observed by changes in intrinsic fluorescence (Frieden, 1990; Touchette et al., 1986; Kuwajima et al., 1991; Jennings et al., 1993). The rate constants for these phases decrease with temperature and increase with lower final urea concentration. The relative amplitudes of the phases observed also vary with temperature and denaturant concentration (Hoeltzli and Frieden, unpublished observation). The amplitude of the first (fastest) phase increases at lower final urea concentration, such that while the first phase represents about half the total amplitude change at 22 °C and a final urea concentration of 1.4 M, it represents only 10% of the amplitude change in these experiments at 5 °C and a final urea concentration of 2.3 M (Table 1). The second phase ($t_{1/2} \sim 5$ s) represents the largest amplitude change under these conditions. The amplitude of the previously observed third phase (k_3) is quite small (~4% at 5° and a final urea concentration of 1.4 M), and it is probably unresolvable under the conditions used in this study. The observed variation in rates and amplitudes suggests that the equilibria between states have been affected, but similar processes to those previously observed are occurring under the conditions used in this study.

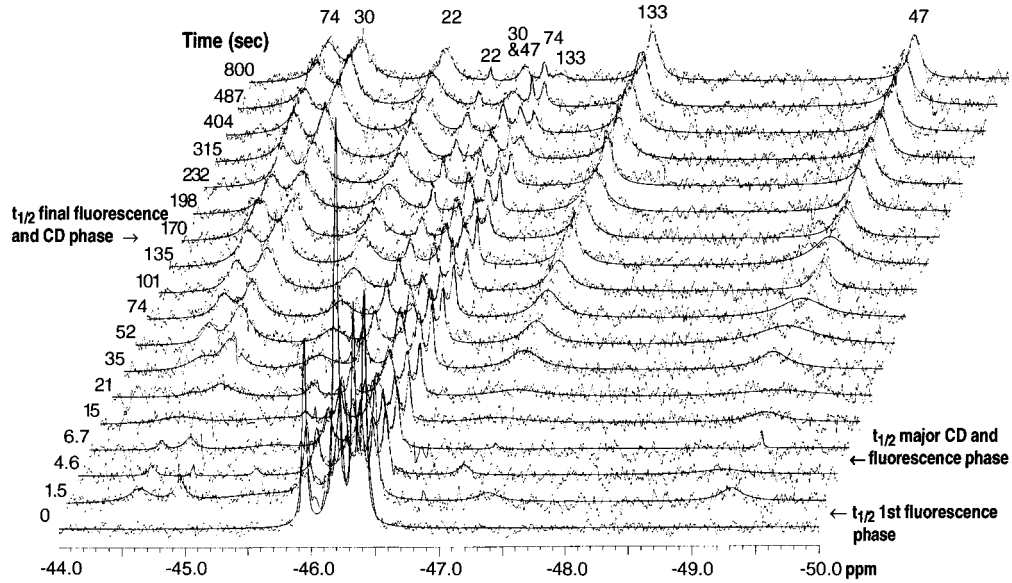

FIGURE 1: Stopped-flow ^{19}F NMR spectra of refolding of 6-^{19}F-tryptophan labeled DHFR on dilution from 4.6 to 2.3 M urea at 5 °C. The time of the data collection is shown on the *z*-axis. The final protein concentration was 0.65 mM. The buffer used contained 50 mM potassium phosphate, pH 7.2, 100 mM KCl, 0.1 mM EDTA, and 15 mM DTT. Other experimental conditions are described in Materials and Methods. For each injection, 100 time points, each a single free induction decay, were obtained. Sixty-five separate injections were summed for each time point and Fourier transformed with a line broadening of 5 Hz. The smooth line represents the area of each resonance, determined using a Lorentzian curve fitting routine (VNMR software, Varian Associates, Palo Alto, CA). The bottom spectrum represents 65 transients of 0.65 mM apo DHFR in 4.6 M urea. Because the first 20 spectra were collected with a recycle time $<4\times$, the resonance intensities between 4.6 and 21 s are expected (and appear) to decrease. At the current signal-to-noise ratio, it cannot be determined whether or not the intensities decrease to a greater extent than predicted for the recycle time used by the measured T_1. As discussed in the text, the native resonances at early times (less than 20 s) most likely represent native protein remaining from a previous injection.

Similar rates and amplitudes are observed for refolding wild-type unlabeled and 6-^{19}F-Trp labeled DHFR.[2] The intrinsic fluorescence of the native and of the unfolded 6-^{19}F-Trp labeled protein is relatively unaffected by urea concentration, making the extent of the fluorescence change during refolding easier to assess. Under the conditions used in this study (5 °C, refolding from 4.6 to 2.3 M urea), there is little (<5%) or no "burst" phase observed either by fluorescence or CD spectroscopies.

The Unfolded Resonances. Figure 1 shows the ^{19}F NMR spectra of 6-^{19}F-tryptophan labeled DHFR at 18 of the 100 time points monitored in real time during refolding on dilution of the protein from 4.6 to 2.3 M urea at 5 °C. The spectrum at 0 time represents an equivalent concentration of protein at equilibrium in 4.6 M urea. Four resonances assigned to the five individual tryptophans in the unfolded state can be resolved. At the first time point observed (1.5 s after initiation of refolding), the intensity of each of these four resonances has been reduced by approximately 50%, but a substantial amount of unfolded protein remains.

The decrease in areas obtained by Lorentzian deconvolution of each resonance assigned to an individual tryptophan in an unfolded environment as a function of time is shown in Figure 2. Each of the four resonances assigned to the five individual tryptophans can be fit by a two exponential function with an initial, missing amplitude of about 50% (Table 1). Within experimental error, all four resonances disappear with similar rates, indicating that all regions of the protein emerge from an unfolded environment at the same time. The data also correspond reasonably well to the rates and amplitudes observed by fluorescence and circular dichroism spectroscopy (see below).

The Native Resonances. Substantial intensity is clearly missing during the first 20 s of the refolding process (Figure 1). This missing intensity may be explained by either chemical shift heterogeneity or by exchange at an intermediate rate between two or more environments. Either explanation indicates that the protein is refolding via one or more intermediates in which the side chains are in a non-native environment which differs substantially from the unfolded state.

The broad lines and poor signal to noise of the native resonances make interpretation of the data obtained at early times problematic. Nevertheless, some tentative conclusions may be drawn from a preliminary analysis of the full time course for the appearance of the native resonances. Figure 3 shows the time dependence of the heights obtained by Lorentzian deconvolution for each resonance assigned to an individual tryptophan in its native environment. Each is well fit by a single exponential function (Table 2, first column). The rates are slow ($t_{1/2} \sim 100$ s), indicating that formation of stable, native side chain environment is a late step in the folding process. The fact that the rates are similar for each tryptophan[3] indicates that all regions of the protein appear to form stable, native side chain environment at about the same time.

In the absence of relaxation effects, conformational exchange, or chemical shift heterogeneity, the ratio of resonance height to resonance area should remain constant. Thus, rates calculated by an exponential fit to resonance areas should correspond to rates calculated by an exponential fit to resonance heights. While this appears to be true for the

[2] In contrast to the wild type protein, the fluorescence of the 6-^{19}F-Trp labeled protein increases for all phases observed upon refolding.

[3] The rate for Trp30 is faster by a factor of 2, indicating that stable, native side chain environment may form slightly faster in this region of the protein. We cannot exclude the possibility that this difference represents experimental error.

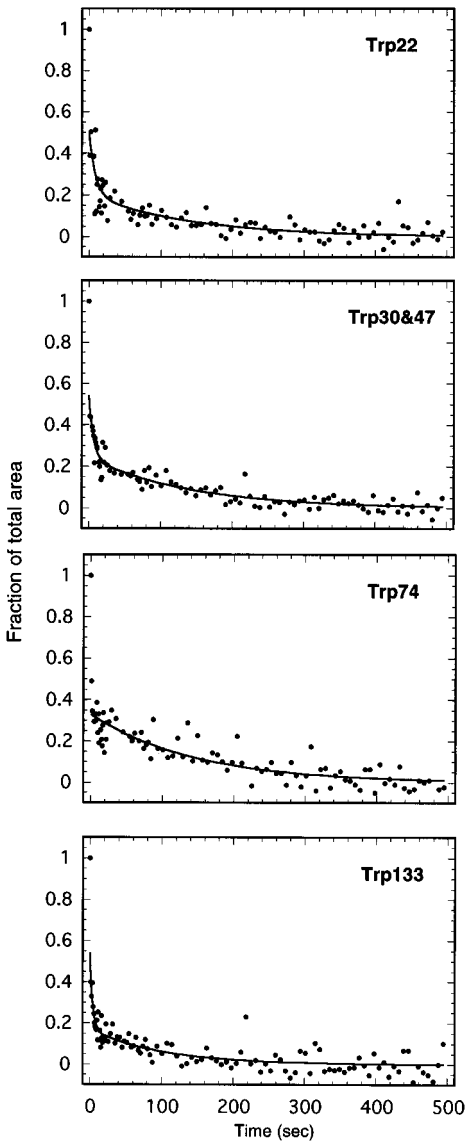

FIGURE 2: Time dependence of the area of the four resonances assigned to Trp22, Trps30 and -47, Trp74, and Trp133 in the unfolded state. Data were collected and resonance areas determined as described in Materials and Methods and the legend for Figure 1. The area of each resonance was normalized to a total area change of 1. The data were fit to a two exponential function (—) to yield the rates and amplitudes reported in Table 1. A three exponential function did not improve the fit. Because the resonance assigned to Trp74 is least well resolved, only one of the two phases present could be fit to an exponential function. An initial phase representing approximately 50% of the total area occurred too rapidly to be observed by this method.

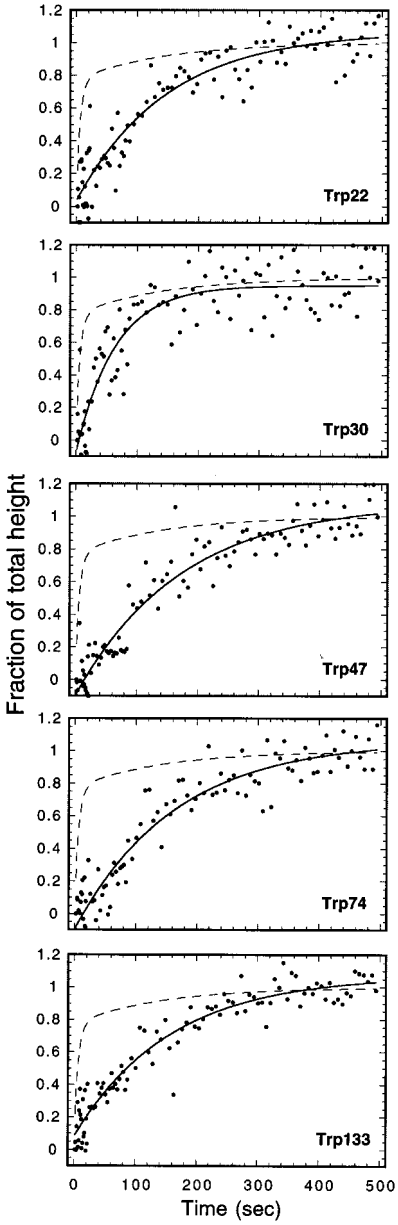

FIGURE 3: Time dependence of the height of the five resonances assigned to the native resonances of Trp22, Trp30, Trp47, Trp74, and Trp133. Data were collected and resonance heights determined as described in Materials and Methods and the legend to Figure 1. The height of each resonance was normalized to a total height change of 1. The data were fit to a single exponential function (—) to yield the rates and amplitudes reported in Table 2. A two exponential function did not improve the fit. The dashed line (- - -) represents the rates and the amplitudes of a three exponential function fit to fluorescence data under identical conditions (Table 1).

resonances of Trp22 and -74, it is apparently not true for those of Trp30, -47 and -133.

Figure 4 shows the areas obtained by Lorentzian deconvolution of the native tryptophan resonances. Each of the resonances corresponding to an individual tryptophan can be well-fit by a single exponential function (Table 2, second column). While the rates obtained by analysis of the resonance areas of Trp22 and -74 correspond with those obtained by analyzing their heights, the rates obtained for Trps30, -47, and -133 appear to be faster than those obtained from the analysis of their heights ($t_{1/2} \sim 14-17$ s vs $t_{1/2} \sim 100$ s). These differences indicate that a process such as conformational exchange or chemical shift heterogeneity is affecting the linewidths of the Trp30, -47, and -133 resonances. This result suggests that regions of the protein around Trps30, -47, and -133 may form native-like tertiary contacts earlier than regions around Trps22 and -74 but that the native-like contacts formed are either unstable and exchange between native and non-native forms or reflect a distribution of heterogeneous side chain environments within a native-like overall structure. Thus, while all regions of the protein appear to form stable, native side chain environment at about the same rate, side chains in some regions appear to sample native-like environments earlier than others.

At 1.5 s, some intensity, perhaps 10−15% of equilibrium area, can be detected at the chemical shifts assigned to the five individual tryptophans in the native state. Under the

Table 2: Appearance of Native Resonances[a]

	using heights		using areas	
	A^b	k (s^{-1})	A^b	k (s^{-1})
Trp22	1.04	0.006	0.72	0.007
Trp30	1.03	0.016	1.02	0.060
Trp47	1.19	0.006	1.14	0.048
Trp74	1.16	0.006	0.76	0.007
Trp133	0.99	0.006	0.98	0.043

[a] Estimated experimental and fitting error of about 20%. [b] Data were fit to a single exponential equation using the program Kaleidagraph (Synergy Software, Reading, PA). Amplitude is given as fraction of total change during the refolding reaction.

experimental conditions used, the injection volume was not sufficient to achieve complete exchange of the reacted material in the cell (see Materials and Methods). This resonance intensity could represent either folded material remaining from the previous injection or a rapid folding population of native molecules. Preliminary ligand binding experiments using stopped-flow fluorescence under the same conditions, but with complete exchange of the cell, show no rapid recovery of ligand binding (Hoeltzli & Frieden, unpublished observations). Thus the presence of the small native peaks at early times most likely represents folded protein remaining from the previous injection.

Fluorescence and Circular Dichroism. To further characterize the environment of the intermediate or intermediates, we examined the changes in intrinsic fluorescence and in circular dichroism at 222 nm under conditions identical to those of the stopped-flow NMR experiment. Data and results of a stopped-flow circular dichroism experiment and results of a stopped-flow fluorescence experiment using identical conditions are shown in Figure 5. By fluorescence, three phases with half-times of ~1, ~5, and ~115 s are observed (Table 1). This result indicates the presence of at least three processes. Although a third exponential does not improve the fit to the circular dichroism data, the fluorescence rates and amplitudes appear to correlate well, indicating that the same processes are being monitored by both techniques.

DISCUSSION

In this report, we have utilized stopped-flow ^{19}F NMR spectroscopy to examine protein folding in real time. We are able for the first time to follow both the loss of unfolded environment and the appearance of native side chain environment. By so doing, we have obtained direct evidence about side chain behavior during protein folding.

The results lead to several major conclusions about protein refolding. First, it is not necessary that all the unfolded protein form a collapsed intermediate in the earliest refolding step. Rather, the data suggest an equilibrium between an early intermediate and the unfolded form. Under the conditions used in these experiments, the ratio of that intermediate to the unfolded from is about 1. It is likely, however, that this equilibrium is affected by both temperature and the magnitude of the denaturant change. As measured by area changes at longer times, all tryptophans leave the unfolded environment at the same rate, suggesting that formation of at least one intermediate is highly cooperative. In this regard, we find that under these conditions there is little or no extremely rapid (less than 5 ms) phase in the fluorescence or CD experiments. This suggests that we are

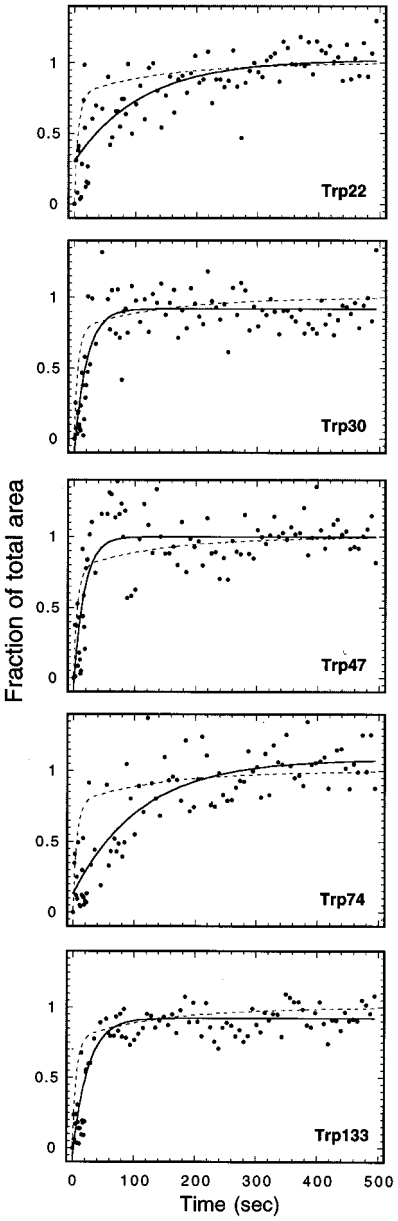

FIGURE 4: Time dependence of the area of the five native resonances assigned to Trp22, Trp30, Trp47, Trp74, and Trp133. Data were collected and resonance areas determined as described in Materials and Methods and the legend to Figure 1. The area of each resonance was normalized to a total area change of 1. The data were fit to a single exponential function (—) to yield the rates and amplitudes reported in Table 2. A two exponential function did not improve the fit. The dashed line (- - -) represents the rates and the amplitudes of a three exponential function fit to fluorescence data under identical conditions (Table 1).

observing the early events in the folding process. Second, since little native resonance is present at a time when 90% of the unfolded resonances have already disappeared, the refolding must involve the formation of an intermediate (or intermediates) in which the side chains either possess considerable mobility or are distributed among a number of non-native environments. The lack of any major resonances at chemical shift values other than those of the unfolded or native protein suggests that no significant population of incorrectly folded stable side chain structures exists at early stages of refolding. Third, major secondary structure formation (as measured by CD changes) and altered solvation states (as measured by fluorescence changes) occur well before

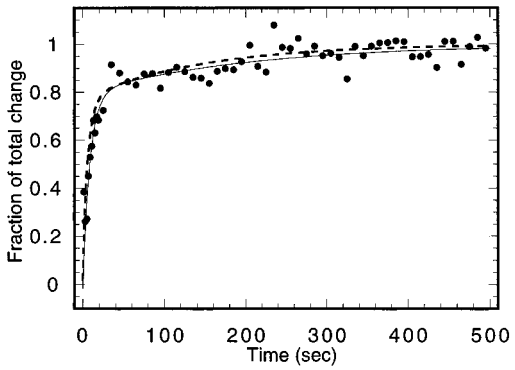

FIGURE 5: Changes in circular dichroism at 222 nm observed on refolding of 6-^{19}F-tryptophan labeled DHFR on dilution from 4.6 to 2.3 M urea at 5 °C. Final protein concentration was 20 μM. The buffer contained 50 mM potassium phosphate, pH 7.2, 100 mM KCl, 0.1 mM EDTA, and 1 mM DTT. Other experimental conditions are described in Materials and Methods. Only 5% of the data points are shown. The circular dichroism data (●) were fit to a two exponential function (—) to yield the rates and amplitudes reported in Table 1. The dashed line (- - -) represents the rates and amplitudes of a three exponential function fit to fluorescence data under identical conditions (final protein concentration 5 μM, Table 1).

stabilization of side chains into their native environment. Finally, side chain stabilization into the native structure is the last step in the refolding process.

Information presented here can also be applied to the specific case of dihydrofolate reductase. While analysis of native resonance heights suggests that formation of stable native side chain environment is cooperative, preliminary analysis of native resonance areas suggests that the regions of the protein surrounding Trps30, -47, and -133 begin to sample unstable native-like side chain environments earlier than Trps22 and -74.

Examination of the crystal structure (Figure 6) in the light of previous results reveals a plausible structural basis for this interpretation. Trp22 is located adjacent to a loop known to be disordered in the apo and binary $NADP^+$ structures (Bystroff et al., 1990; Bystroff & Kraut, 1991). While Trp74 is in a β-strand, it is located in the adenine-binding domain near the surface of the protein. Previous work in this laboratory (Frieden, 1990) revealed that the ligand binding sites form at different rates, with the dihydrofolate-binding site forming relatively early, while the $NADP^+$ binding site forms late in the folding process.

A study of DHFR folding as assessed by hydrogen−deuterium exchange by Jones and Matthews (1995) reveals that 6 of 8 β-strands show early protection. Trps30 and -133 flank this region of the protein, and early sampling of native-like environments by these residues is consistent with the exchange data. In addition, Hall and Frieden (1989) showed that a peptide fragment from this same region (Gln102-Glu154) was most effective in inhibiting recovery of native protein, suggesting that this region was crucial to the folding of the protein.

The suggestion that native structure in the region around Trp47 forms early while native structure in the region around Trp74 forms late would appear to contradict the interpretation of data obtained by Matthews and co-workers. Based on CD and fluorescence data obtained with site-directed mutants of Trp74, these investigators proposed that an early folding step was a hydrophobic collapse of these two tryptophan residues into a native-like environment (Garvey et al., 1989; Kuwajima et al., 1991). While fluorescence and CD data clearly indicate an effect of mutating Trp74 upon the first phase, the structural interpretation of this effect is less certain. The amides surrounding this region do not demonstrate the earliest protection from exchange (Jones & Matthews, 1995).

The unfolded resonances disappear in at least three phases, indicating that at least three processes are involved in the loss of unfolded environment. However, the presence of three processes does not require the presence of three slowly-exchanging populations of unfolded protein in order to create an adequate kinetic model. The simplest kinetic model to simulate the unfolded resonance intensity data, the fluorescence and CD data, and the major phases of the native area and height data is:

$$U \rightleftharpoons I1 \rightleftharpoons I2 \rightarrow N$$

In order to adequately model the data, both of the first two steps must be reversible. The shallow urea jump, high final concentration of urea, and low temperature used in this folding study probably maximize reversibility. In this model, I2 would represent the native-like intermediate more rapidly

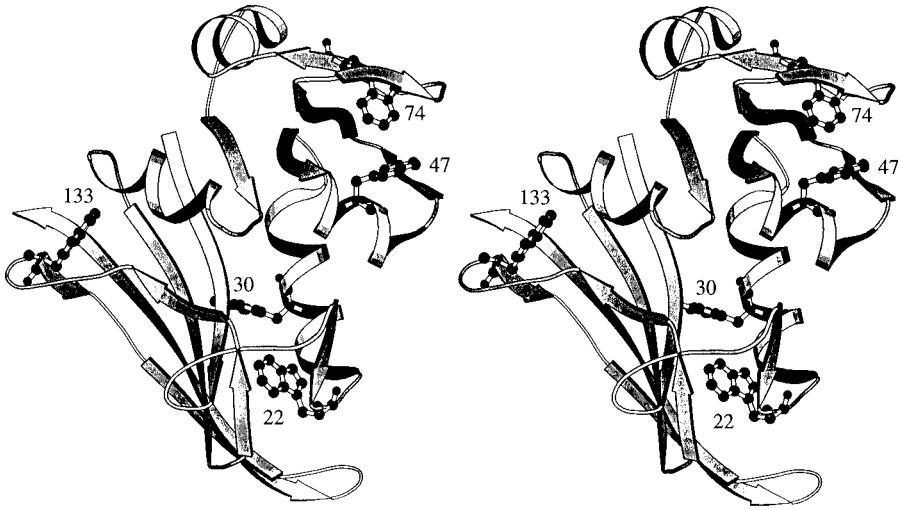

FIGURE 6: Stereo ribbon diagram of the ternary complex (Bystroff et al., 1990) of *E. coli* DHFR drawn using the program MOLSCRIPT (Kraulis, 1991). Ligands are not shown. The side chains of the 5 tryptophan residues at positions 22, 30, 47, 74, and 133 are represented in ball-and-stick form.

populated by Trps30, -47 and -133 while, for Trps22 and -74, the formation of I2 is rate limiting. All 5 tryptophans form N at the same rate and in this sense the final step is cooperative.

Early in the folding reaction, some intensity (perhaps 10−15% of equilibrium area) can be detected at the chemical shifts assigned to the five individual tryptophans in the native state. This intensity could either represent a rapid folding population of native molecules, or folded material remaining from the previous injection. If an early folding population of native or native-like molecules exists, this model is not able to adequately simulate it. Because preliminary ligand binding studies performed under the same conditions do not indicate the existence of a rapidly-forming population able to bind ligand, we believe that this resonance intensity most likely represents folded material remaining from the previous injection.

It should be noted that there is seldom a unique mechanism able to model a set of kinetic data. This is perhaps especially true when the physical interpretation of amplitude changes is uncertain, as is often the case with observed fluorescence changes. Depending upon the assumptions one makes, the data in this study can also be adequately simulated by several other models, including a three-channel model in which at least two of the pathways contain intermediates and a model involving at least two off-pathway intermediates.

A number of mechanisms have previously been proposed to explain the folding of dihydrofolate reductase (Frieden, 1990; Touchette et al., 1986; Kuwajima et al., 1991; Jennings et al., 1993; Jones et al., 1995). Our results do not exclude any previously suggested mechanism. We suggest the above mechanism only as the simplest adequate model. A reversible sequential model is adequate to explain hydrogen−deuterium exchange data showing early protection of only 50% of the molecules (Jones & Matthews, 1995).

In a previous study (Hoeltzli & Frieden, 1995), we utilized stopped-flow ^{19}F NMR spectroscopy to study the unfolding of DHFR in real time. We observed the rapid formation of an intermediate containing substantial secondary structure and native-like fluorescence, but possessing altered side chain mobility. A similar observation of an unfolding intermediate containing secondary structure and protection from hydrogen exchange, but possessing altered side chain mobility, was made for ribonuclease A (Kiefhaber & Baldwin, 1995; Kiefhaber et al., 1995) using NMR spectroscopy in real time. We suggested that the dry molten globule postulated by Shakhnovich and Finkelstein (1989) might explain the observation that this intermediate apparently possesses native-like fluorescence.

In a recent study of refolding bovine α-lactalbumin (in the absence of a structural calcium ion) using a real-time NMR technique similar to that for ribonuclease A (Balbach et al., 1995), formation of secondary structure (as assessed by far-UV CD) preceded the major change in fluorescence intensity, while the latter correlated with appearance of native side chain intensity. These results differ from the data presented here in that, for DHFR, major changes in fluorescence and far-UV CD occurred simultaneously and well before the formation of stable, native side chain environment. While formation of side chain environment in bovine α-lactalbumin was also observed to be cooperative, and to be a rate-limiting step, the folding intermediate formed in α-lactalbumin apparently differs in solvent accessibility.

It is interesting that, in our study, the major changes in fluorescence appear to correspond to the loss of the unfolded environment rather than to formation of stable native side chain structure. While changes in fluorescence intensity have been attributed to formation of stable, native tertiary structure, this observation has often been made in proteins lacking a structural cofactor present in the intact native protein and has not been confirmed by other methods of evaluating the native qualities of the structure. It has previously been demonstrated that recovery of enzymatic activity occurs after the major fluorescence change observed in the folding of DHFR and is a late step in the folding of this protein (Frieden, 1990). Since few proteins have been examined by a probe of tertiary structure (recovery of enzymatic activity or native side chain chemical shifts and relaxation properties) in addition to fluorescence, it remains to be seen how general this observation may prove.

In this refolding study, we observe a folding intermediate or intermediates with properties very similar to those observed on the unfolding pathway: a molten globule-like structure possessing altered side chain mobility, substantial secondary structure, and native-like fluorescence. If the observation of structures containing native-like secondary structure and solvent protection (as evidenced by fluorescence), but non-native side chain behavior, proves general, it would provide evidence to support the theory that solvent exclusion serves as a crucial driving force for protein folding and would suggest that formation of native side chain contacts might be a general rate-limiting step in protein folding.

ACKNOWLEDGMENT

We thank Dr. D. A. D'Avignon for help with the experimental NMR work, Drs. D. P. Cistola and K. B. Hall for the use of the proton/fluorine probe, Ms. Maureen Zetlmeisl for excellent technical assistance in protein purification, and Dr. A. C. Clark for helpful discussions.

REFERENCES

Baccanari, D., Phillips, A., Smith, S., Sinski, D., & Burchall, J. (1975) *Biochemistry 14*, 5267−5273.
Balbach, J., Forge, V., van Nuland, N. A. J., Winder, S. L., Hore, P. J., & Dobson, C. M. (1995) *Nature, Struct. Biol. 2*, 865−870.
Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 84−91.
Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., & Kraut, J. (1982) *J. Biol. Chem. 257*, 13650−13662.
Bystroff, C., & Kraut, J. (1991) *Biochemistry 30*, 2227−2239.
Bystroff, C., Oatley, S. J., & Kraut, J. (1990) *Biochemistry 29*, 3263−3277.
Englander, S. W., & Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 243−265.
Fierke, C. A., Johnson, K. A., & Benkovic, S. J. (1987) *Biochemistry 26*, 4085−4092.
Frieden, C. (1990) *Proc. Natl. Acad.Sci. U.S.A. 87*, 4413−4416.
Garvey, E. P., Swank, J., & Matthews, C. R. (1989) *Proteins 6*, 259−266.
Hall, J. G., & Frieden, C. (1989) *Proc. Natl. Acad.Sci. U.S.A. 86*, 3060−3064.
Hoeltzli, S. D., & Frieden, C. (1994) *Biochemistry 33*, 5502−5509.
Hoeltzli, S. D., & Frieden, C. (1995) *Proc. Natl. Acad.Sci. U.S.A. 92*, 9318−9322.
Hoeltzli, S. D., Ropson, I. J., & Frieden, C. (1994) *Tech. Protein Chem. V*, 455−466.
Jennings, P. A., Finn, B. E., Jones, B. E., & Matthews, C. R. (1993) *Biochemistry 32*, 3783−3789.
Jones, B. E., & Matthews, C. R. (1995) *Protein Sci. 4*, 167−177.

Jones, B. E., Beechem, J. M., & Matthews, C. R. (1995) *Biochemistry 34*, 1867−1877.
Kiefhaber, T., & Baldwin, R. L. (1995) *Proc. Natl. Acad.Sci. U.S.A. 92*, 2657−2661.
Kiefhaber, T., Labhardt, A. M., & Baldwin, R. L. (1995) *Nature 375*, 513−515.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R., & Sugai, S. (1991) *Biochemistry 30*, 7693−7703.
Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
Penner, M. H., & Frieden, C. (1987) *J. Biol. Chem. 262*, 15908−15914.
Roder, H. (1989) *Methods Enzymol. 176*, 446−473.
Shakhnovich, E. I., & Finkelstein, A. V. (1989) *Biopolymers 28*, 1667−1680.
Touchette, N. A., Perry, K. M., & Matthews, C. R. (1986) *Biochemistry 25*, 5445−5452.

BI961896G