Structure of a Hydrophobically Collapsed Intermediate on the Conformational Folding Pathway of Ribonuclease A Probed by Hydrogen−Deuterium Exchange†

Walid A. Houry and Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

Received May 7, 1996; Revised Manuscript Received July 11, 1996⊗

ABSTRACT: The unfolded state of disulfide-intact bovine pancreatic ribonuclease A is a heterogeneous mixture of unfolded species which have different X−Pro peptide bond conformations. One of these unfolded species, labeled U_{vf}, has all its X−Pro peptide bonds in the native conformation. Therefore, the refolding of U_{vf} is a purely conformational folding process which is not complicated by cis−trans X−Pro peptide bond isomerization. There are two identifiable intermediates on the folding pathway of U_{vf}: one which is a largely unfolded intermediate (I_U) and another which is a hydrophobically collapsed intermediate (I_Φ). An instrument was built, and experiments were designed to study the structure in I_U and I_Φ by hydrogen−deuterium exchange. These experiments are a combination of a double-jump experiment followed by a pulse-labeling experiment. The native protein was first unfolded to populate U_{vf} to more than 99%, and then U_{vf} was refolded for a specified period of time. After refolding, hydrogen−deuterium exchange of the backbone amides was initiated for a given time by raising the pH. Subsequently, the exchange was quenched and the protein was allowed to continue to fold to the native state. The extent of exchange was determined quantitatively by two-dimensional NMR spectroscopy. The data indicate that I_U has no secondary structure that can protect the backbone amides from exchange under the conditions employed. On the other hand, in I_Φ, the second helix (residues 24−34) and a large part of the β-sheet region of the protein are formed, while the rest of the protein molecule remains unstructured. In general, the protection factors in I_Φ are low, indicating that this intermediate has a dynamic structure. Our observations are consistent with I_Φ being a molten-globule-like intermediate. The regular structure formed in I_Φ is much less than that observed in a hydrogen-bonded intermediate (I_1) populated early on the major slow-refolding pathway of the protein [Udgaonkar, J. B., & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87,* 8197−8201]; in addition, the structure in I_Φ has much lower stability than that in I_1. This implies that a slower refolding rate allows for a higher cooperativity between the different structural elements of the protein, resulting in the formation of more stable (native-like) intermediates (as in I_1) during the folding process.

The protein folding pathway is the kinetic process that best describes the events that occur when an unfolded polypeptide chain, which has no regular structure, proceeds to fold to its native state which is characterized by a highly ordered structure. The interactions that govern this process are strictly those defined by the amino acid sequence of the protein (Anfinsen, 1973). An unfolded chain cannot attain its native state by some random search of the conformational space because of the extensive time required for such a search (Levinthal, 1969). Therefore, the initial events in protein folding must restrict the conformational space and must specify the pathway that the chain should follow to reach its native state. Hence, it is important to characterize the early intermediates formed on the folding pathway of a protein. Investigating the structures of these intermediates would help identify the nature of the interactions that determine the folding pathway of the protein. These interactions, consequently, predetermine the structure of the native state of the polypeptide chain.

In an attempt to identify the nature of the earliest intermediates formed on the folding pathway of a protein, a pulse labeling technique has recently been introduced (Udgaonkar & Baldwin, 1988; Roder et al., 1988). This technique utilizes hydrogen−deuterium exchange to identify the backbone amides that become protected during the folding process of the protein. The pulse-labeling technique consists basically of three steps: a refolding step followed by a labeling step which is then followed by a quenching step. It is usually carried out as follows. The unfolded completely deuterated protein is allowed to fold for some specified time at low pH where the exchange is essentially prevented. Then, the protein is diluted into H_2O at high pH to initiate exchange. The exchange is then quenched after some determined time, and the protein is allowed to proceed to the native state. Those backbone amides that are protected during the folding process will not exchange and will remain deuterated, while those not protected will become protonated. The extent of exchange is evaluated quantitatively by measuring the amide peak heights in a two-dimensional NMR spectrum.

The pulse-labeling technique has been used to study the folding pathway of several proteins (Bycroft et al., 1990; Lu & Dahlquist, 1992; Briggs & Roder, 1992; Radford et al., 1992; Varley et al., 1993; Mullins et al., 1993; Jennings

† This work was supported by Grant GM-24893 from the National Institute of General Medical Sciences of the National Institutes of Health. Support was also received from the National Foundation for Cancer Research.

* Author to whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1996.

S0006-2960(96)01085-9 CCC: $12.00

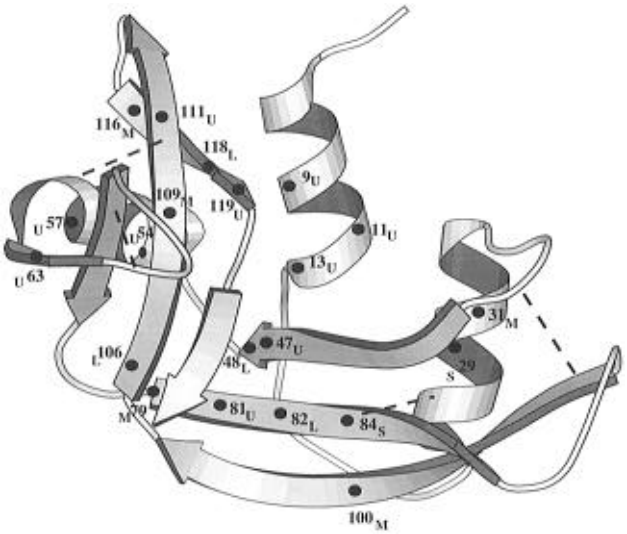


FIGURE 1: Ribbon representation of the structure of RNase A drawn with the program MOLSCRIPT (Kraulis, 1991). The coordinates used are those given in the file 7RSA of the Brookhaven Protein Data Bank. The heavy dots give the positions of the 21 amino acid residues which were used to monitor structure formation during the refolding of U_{vf}. The residue number is written next to each dot. The subscript on each residue number refers to the level of protection observed in I_Φ for that given residue (refer to Results and Discussion). Amino acid residues labeled U have protection factors of less than 1.5 in I_Φ ($P < 1.5$), those labeled L protection factors between 1.5 and 5 in I_Φ ($1.5 < P < 5$), those labeled M protection factors between 5 and 50 in I_Φ ($5 < P < 50$), while those labeled S protection factors greater than 100 in I_Φ ($P > 100$). There are four disulfides in RNase A: 26−84, 40−95, 58−110, and 65−72. For simplicity, the disulfides are shown as dashed lines between the C^α atoms of the Cys residues.

& Wright, 1993; Jacobs & Fox, 1994; Jones & Matthews, 1995), including disulfide-intact bovine pancreatic ribonuclease A (RNase A)[1] (Udgaonkar & Baldwin, 1988, 1990, 1995), the protein of interest here. The structure of RNase A (Wlodawer et al., 1988) consists of three helices and a large β-sheet region (Figure 1). The first helix is at the N-terminus spanning residues 3−13. The second helix directly follows the first helix, and it consists of residues 24−34. Then there is a small β-sheet region consisting of residues 41−48, which is followed by the third helix that is made up of residues 50−60. After the third helix, there is a loop region which spans the 65−72 disulfide bond. The rest of the protein molecule is made up of antiparallel β-sheet regions in which there are two type VI β-turns containing cis X−Pro 93 and 114 peptide bonds. There are two other prolines in the protein which have trans X−Pro peptide bonds, namely X−Pro 42 and 117.

The refolding kinetics of RNase A are complicated by the heterogeneity of the unfolded state which is characterized by the presence of multiple unfolded species. The heterogeneity of the unfolded state of the protein arises mainly from X−Pro peptide bond isomerization. This was first suggested by Brandts et al. (1975). When the protein unfolds, the X−Pro peptide bonds undergo cis−trans isomerization to populate both the cis and trans conformations. The trans conformation is the more favored one in the unfolded state. The suggestion of Brandts et al. (1975) has recently been supported by studies of mutants of RNase A (Schultz & Baldwin, 1992; Schultz et al., 1992; Dodge et al., 1994; Dodge & Scheraga, 1996) in which alanines, glycines, or serines were substituted for the prolines.

Five different phases have been observed experimentally for the refolding of RNase A (Dodge & Scheraga, 1996; Houry & Scheraga, 1996). These five phases correspond to the refolding of (at least) five different unfolded species present in the unfolded state of the protein. These species are U_{vf}, U_f, U_m, U_s^{II}, and U_s^{I}, the very-fast-, fast-, medium-, major slow-, and minor slow-refolding species, respectively. All the X−Pro peptide bonds of U_{vf} are in their native conformation, and hence, the refolding of U_{vf} is a purely conformational folding reaction (Houry et al., 1995). The other unfolded species have one or more nonnative X−Pro peptide bonds. Udgaonkar and Baldwin (1988, 1990, 1995) have used the pulse-labeling technique to study the structure of an early-folding intermediate (I_1) on the folding pathway of the major slow-folding species (U_s^{II}). In the current investigation, we study the folding of U_{vf} by this technique.

The difficulty in studying the refolding kinetics of U_{vf} arises from the fact that the population of U_{vf} is only 2−5% in the equilibrium unfolded state of the protein (Houry et al., 1994; Dodge & Scheraga, 1996; Houry & Scheraga, 1996). Therefore, to overcome this problem, a double-jump technique was employed. In this technique, the native protein is unfolded for a time period long enough to form >99% U_{vf} but short enough so as not to form the other unfolded species. Then U_{vf} is refolded, and its refolding process is monitored by different methods. If the double-jump experiment is coupled to the pulse-labeling experiment, then we should be able to monitor the refolding process of U_{vf} by hydrogen−deuterium exchange. We call such an experiment a DJ−HD experiment. The procedure for this experiment is described in Materials and Methods.

Two intermediates have previously been detected on the refolding pathway of U_{vf} (Houry et al., 1995, 1996): a largely unfolded intermediate (I_U) and a hydrophobically collapsed intermediate (I_Φ). The DJ−HD experiments have enabled us to obtain a detailed picture of the structure of these two intermediates. The results obtained are consistent with hydrophobic collapse as the initial event in protein folding.

MATERIALS AND METHODS

Reagents. Acetic acid (HAc), HCl, and NaOH were purchased from Fisher. Glycine (Gly), DCl, and NaOD were purchased from Sigma. D_2O was purchased from Isotec Inc. or Cambridge Isotope Laboratories, while deuterated acetic acid (CD_3COOD) was purchased from Cambridge Isotope Laboratories. Citric acid was purchased from Pierce Chemical Co., and formic acid from Fluka. Ultrapure guanidine hydrochloride was purchased from ICN Biochemicals. GdnHCl concentrations were determined by refractive index (Nozaki, 1972).

Protein Purification. RNase A, types I-A and II-A, was purchased from Sigma and was purified further by cation-

[1] Abbreviations: RNase A, disulfide-intact bovine pancreatic ribonuclease A; DJ−HD experiment, an experiment which is a combination of a double-jump experiment followed by a pulse-labeling experiment; HD exchange, hydrogen−deuterium exchange; NH or ND, backbone amide proton or deuterium; GdnHCl, guanidine hydrochloride; DQF-COSY, two-dimensional double-quantum-filtered correlation spectroscopy; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; TOCSY, two-dimensional total correlation spectroscopy; amide peak heights, the peak heights obtained from DQF-COSY experiments, processed in absolute-value mode, of the NH−C^αH cross-peaks referenced to the Tyr 25 $C^\delta HH'-C^\epsilon HH'$ cross-peak.

exchange chromatography according to the procedure of Rothwarf and Scheraga (1993). The purity of the protein was checked by using a Hydropore-5-SCX column (Rainin) on an 8700 SpectraPhysics HPLC apparatus and was found to be >99% pure.

Instrument Used for the DJ−HD Experiments. The instrument used for the DJ−HD experiments was constructed in the Physics Machine Shop at Cornell University. The design of the instrument is based largely on that of the Hi-Tech Scientific PQ/SF-53 stopped flow device. It uses pneumatic rams to drive rapidly up to eight different syringes containing different buffers, thereby allowing for the rapid mixing of these different buffers. The instrument consists of three main systems: a liquid flow system that directs the flow of the buffers/protein solutions through the different mixers and tubings, a gas flow system which directs the flow of the N_2 gas that drives the pneumatic rams, and a circulating bath that maintains the temperature of the whole setup. Several components of these systems were purchased from different vendors.

For the liquid flow system, the syringes were purchased from Hamilton (Reno, NV). Glass-filled Teflon high-pressure distribution valves used for the flow circuit were purchased from Omnifit USA (Toms River, NJ). Inline check valves, fittings, and PEEK tubings were obtained from Upchurch Scientific (Oak Harbor, WA). A four-jet Berger mixer (Berger et al., 1968) from Hi-Tech Scientific (Wiltshire, England) was used to achieve rapid and efficient mixing of the different GdnHCl-containing buffers. The volumes of the delay lines were measured accurately using the Hamilton syringes. A model 230 submersible stirrer was purchased from VWR Scientific (Piscataway, NJ).

For the gas flow system, the electronic solenoid valves were purchased from Cole-Parmer (Vernon Hills, IL) or Clippard (Cincinnati, OH). The pneumatic valves and the pneumatic rams were purchased from Clippard. Miniature air regulators were purchased from Aro Corp. (Bryan, OH), while a main high-pressure air regulator was obtained from Fisher Scientific. Two timers were connected to the instrument to actuate the electronic valves and to measure the delay times. One was a TDU-53 timer obtained from Hi-Tech Scientific, while the other one was constructed by ITL, Inc. (Ithaca, NY).

The temperature of the system was maintained constant with a Forma Scientific, Inc. (Marietta, OH) circulating bath connected to a Little Giant pump (Oklahoma City, OK). The temperature was monitored with an electronic thermometer from Baxter (Edison, NJ). The reading of the thermometer was checked against the reading of an external thermistor (Fisher Scientific). The two readings were found to be similar within 0.2 °C.

Solutions containing only GdnHCl were used to check that the different buffers in the instrument were mixed in the expected ratio. The final GdnHCl concentration after each mixing event was measured by refractive index (Nozaki, 1972) and was found to agree with the expected concentration.

Deuterating RNase A. The protein was deuterated as follows. RNase A was dissolved in 0.7 mL of D_2O at a concentration of about 60 mg/mL, and the pH[2] of the sample was adjusted to 4.0 with DCl or NaOD. Several samples were prepared at the same time. The samples were filtered through a 0.2 μ filter (Gelman). The protein samples were then heated at 60 °C for 10 min and allowed to cool to room temperature. They were subsequently frozen and lyophilized. The entire procedure was then repeated. One-dimensional NMR spectra on a Varian VXR-400S spectrometer showed that all the amide NHs were completely exchanged.

DJ−HD Experiments. The DJ−HD experiments, carried out on the instrument described above, were designed to examine structure formation during the refolding of U_{vf}. The temperature in all of these experiments was 5 °C. Three different conditions were employed.

Experiments were carried out as follows. All the solutions used in these experiments were in H_2O except the initial protein solution which was in D_2O. The completely deuterated protein at a concentration of 51 mg/mL and pH 4.0 in D_2O was unfolded at 4.2 M GdnHCl (pH 2.0) by 1:2.5 dilution with 5.88 M GdnHCl and 50 mM Gly at pH 1.5. The protein was allowed to unfold for 1 s which resulted in the formation of >99% U_{vf} without significant formation of the other unfolded species (Houry et al., 1994, 1995). After 1 s, the protein was refolded for 6 ms at (A) 0.7 M GdnHCl at pH 3.0 or (B) 2.5 M GdnHCl at pH 4.0 or kept unfolded at (C) 4.2 M GdnHCl at pH 2.0 by 1:5 dilution with (A) 0 M GdnHCl and 50 mM citric acid at pH 3.46, (B) 2.16 M GdnHCl and 50 mM formic acid at pH 4.6, or (C) 4.2 M GdnHCl and 50 mM Gly at pH 2.0. The 6 ms is the shortest refolding time that could be achieved in the current setup of the instrument. No (significant) exchange takes place during the unfolding or refolding steps at pH 2.0−4.0 because, under these conditions, the rate of exchange is on the order of 2−100 min (Bai et al., 1993). Both 0.7 M GdnHCl at pH 3.0 and 2.5 M GdnHCl at pH 4.0 at 5 °C are baseline folding conditions (conditions where the native protein is 100% folded) (Houry et al., 1994, 1995). After refolding for 6 ms, the protein was labeled with H_2O at pH 9.0 for 20 ms to exchange the NDs for NHs. This was achieved by diluting the refolded protein in a ratio of 1.2:1 with (A) 0.7 M GdnHCl and 333 mM Gly at pH 9.8, (B) 2.5 M GdnHCl and 333 mM Gly at pH 9.1, or (C) 4.2 M GdnHCl and 333 mM Gly at pH 9.5. The labeling pulse was terminated after 20 ms by lowering the pH to 3.0. This was accomplished by injecting the solution from the labeling pulse, which had a volume of 4.6 mL, into a beaker containing 20 mL of 100 mM acetic acid (pH 2.0) quench buffer while stirring vigorously. The quench buffer in the beaker was also at 5 °C. At the end of the experiment, the final solution in the beaker consisted of 2.0 mg of RNase A and (A) 0.15 M GdnHCl, (B) 0.46 M GdnHCl, or (C) 0.75 M GdnHCl and pH 3.0. Refolding was allowed to go to completion (τ_{vf} under the quench conditions is about 15−25 ms), and then the samples were placed in the cold room for desalting. In order to collect enough protein for the NMR experiments, the above procedure was repeated 10 times for each condition. In other words, 10 samples from the DJ−HD experiments were needed to make up one NMR sample. Two NMR samples were obtained for each condition. The DJ−

[2] All pH values are glass electrode readings with no correction for isotope effects. In this paper, pH is taken to mean pH or pD.

HD experiment can be described schematically as follows:

(a) completely deuterated folded protein in D_2O

↓

(b) unfold for 1 s at 4.2 M GdnHCl and pH 2.0 to form >99% U_{vf}

↓

(c) refold for 6 ms at 0.7 M GdnHCl and pH 3.0, 2.5 M GdnHCl at pH 4.0, or 4.2 M GdnHCl at pH 2.0, to form the different intermediates on the refolding pathway of U_{vf} in addition to the native species

↓

(d) pulse protein with H_2O at pH 9.0 for 20 ms using the same GdnHCl concentration as in (c) to initiate the exchange of the unprotected amides; the protein continues to fold during this step

↓

(e) quench the exchange by lowering the pH to 3.0 and allow the protein to continue to fold to the native state

The percentage of D_2O at each step was as follows: (a) 100%, (b) 29%, (c) 5%, (d) 3%, and (e) 0.2%. A schematic diagram of the instrument used for the above experiments is shown in Figure 2.

Desalting and Concentrating the Samples. The samples obtained from the DJ−HD experiments were collected from the quench beaker and placed in the cold room. They were desalted directly after the end of the experiment. All the desalting steps were carried out at 5 °C to minimize any exchange. The 10 samples were pooled together, resulting in a solution with a total volume of about 250 mL containing about 20 mg of protein. A 400 mL Amicon cell with a YM-10 membrane was initially used to concentrate the protein solution from 250 to about 25 mL. Then the solution was exchanged into D_2O by diafiltration using 100 mL of 100 mM CD_3COOD in D_2O (pD = 2.8). The solution was then placed in a 10 mL Amicon cell with a YM-10 membrane and was further concentrated to a volume of about 2.0 mL. The sample was then transferred into a Centricon-10 device and was further concentrated to a final volume of 0.3−0.4 mL by spinning at 7000 rpm in a Beckman Model J2-21 centrifuge at 5 °C using a JA-20 rotor. Prior to using the Centricon-10 device, it was necessary to remove the glycerol used to preserve the membrane in the device. This was accomplished by spinning the device with 100 mM CD_3COOD in D_2O several times. The protein sample was then filtered through a 0.2 μ filter (Gelman), and its pD was adjusted to 3.2 with DCl and/or NaOD. At the end, the protein concentration in the sample was 2−3 mM as determined by absorbance at 277.5 nm (Sela & Anfinsen, 1957), and the D_2O was >99%. The sample was then stored at −70 °C until analysis by two-dimensional NMR spectroscopy.

The DJ−HD experiments and the subsequent desalting/concentrating of the protein samples were carried out on two consecutive days. It should be pointed out that alternative procedures were also explored to reduce the time required to desalt/concentrate a 250 mL sample to 0.3 mL. We checked whether it is possible to lyophilize the protein after desalting it; however, this resulted in the scrambling of the label. Our observation is consistent with the results of Klibanov and his co-workers, who found that a rearrangement of the secondary structure of the protein takes place upon lyophilization (Desai et al., 1994; Costantino et al., 1995; Griebenow & Klibanov, 1995). Furthermore, we noticed that freezing the protein in a large volume of D_2O or H_2O (>100 mL) also resulted in scrambling of the label. The reason for this behavior is not known to us and might be due to some local pH effects when such a large solution freezes slowly. We did not observe such behavior when the protein sample was frozen in small volumes of D_2O or H_2O buffers.

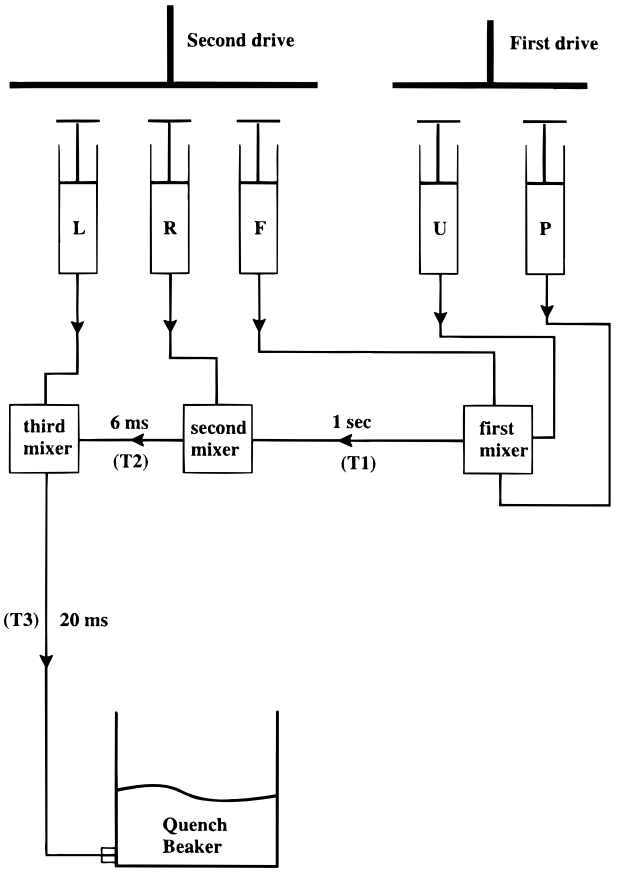


FIGURE 2: Schematic drawing of the instrument used to carry out the DJ−HD experiments. The protein is placed in syringe P, the unfolding buffer in syringe U, the flush buffer in syringe F, the refolding buffer in syringe R, and the labeling buffer in syringe L. Initially, syringes P and U are driven, and the two buffers from P and U meet at the first mixer to unfold the protein. The unfolded protein remains in the tubing T1, between the first mixer and the second mixer, for a time period of about 1 s (the unfolding time), resulting in the formation of >99% U_{vf}. After about 1 s from the end of the first push, syringes F, R, and L are driven together to flush the protein out of tubing T1 and then to refold it at the second mixer and to label it at the third mixer. The protein solution is subsequently injected into the beaker where the exchange is quenched by lowering the pH. The time of travel from the second mixer to the third mixer (in tubing T2) is 6 ms (the refolding time), while the time of travel from the third mixer to the beaker (in tubing T3) is 20 ms. The whole system is thermostated at 5 °C. A description of the buffers used is given in Materials and Methods.

NMR Experiments. The desalted/concentrated sample from the DJ−HD experiments was removed from the −70 °C freezer and was allowed to thaw at room temperature. The sample (2−3 mM in 100 mM CD_3COOD in D_2O at pH 3.2) was then placed in a BMS-005V Shigemi NMR microtube (Shigemi Inc., Allison Park, PA). DQF-COSY (Piantini et al., 1982) spectra were recorded at 500 MHz on a Varian Unity 500 instrument at 25 °C. Sixteen transients of 1024 complex points per transient were obtained for each of 256 t_1 increments. The spectral width was 5999.7 Hz. The spectra were processed in absolute-value mode on a SUN Sparcstation LX using the program Felix version 2.3 (Biosym Technologies, San Diego, CA). Prior to Fourier transformation in each dimension (t_2 and t_1), the data were zero-filled to 2048 points and multiplied by a sine bell-squared window function. The intensities of the NH−C^{α}H cross-peaks were obtained by measuring their peak heights. The peak heights were normalized to the peak height of the nonexchanging Tyr 25 $C^{\delta}HH'$−$C^{\epsilon}HH'$ cross-peak in order to be able to compare samples with different protein concentrations. The cross-peaks in the two-dimensional NMR spectra have been assigned by Robertson et al. (1989) and by Rico et al. (1989, 1993). We further checked the assignments under the condition used for our experiments by carrying out TOCSY (Braunschweiler & Ernst, 1983) and NOESY (Jeener et al., 1979; Macura & Ernst, 1980) experiments at pH 3.2 and 25 °C. The (relative) uncertainty in measuring a given peak height from two repeats was between 5 and 10%.

In this paper, when we refer to the amide peak heights, we mean the peak heights of the NH−C^{α}H cross-peaks after being normalized to the Tyr 25 $C^{\delta}HH'$−$C^{\epsilon}HH'$ cross-peak.

Control Experiments. Several control experiments were carried out to determine the set of backbone amide NHs that could serve as useful probes during the refolding of U_{vf}. These amide hydrogens should be resistant to exchange in the native state, and they should also survive the desalting procedure described above.

In order to determine the amide hydrogens that are resistant to exchange in the native state, the following experiments were carried out. A fully protonated sample of lyophilized RNase A was dissolved in 100% D_2O, and its pH was adjusted to 3.2 with DCl and NaOD. Then, about 3 h after the protein was dissolved in D_2O, a DQF-COSY spectrum was obtained for the sample at 25 °C. The time required to acquire the spectrum was 9 h. The protein was then kept at 25 °C for an additional 16 h (i.e. 28 h from the time that the protein was first dissolved), and then another DQF-COSY spectrum was recorded for the sample. The spectra were processed as described above, and the amide peak heights from the two spectra were obtained. Thirty-six amide NHs were found whose peak height in the second spectrum did not decrease in intensity by more than 10−15%.

From the 36 amides found to be resistant to exchange in the native state, we wanted to select those that are also resistant to exchange during the desalting/concentrating procedure described above. During desalting/concentrating, the samples are initially placed in a 400 mL Amicon cell, then a 10 mL Amicon cell, and then a Centricon-10 device. Consequently, three control experiments were carried out where the desalting/concentrating was started in the 400 mL Amicon cell, the 10 mL Amicon cell, or the Centricon-10 device and then proceeding as described above. DQF-COSY spectra were recorded for the three control samples, and the amide peak heights that were within 10−15% of the peak heights obtained for the sample that had been prepared by dissolving fully protonated RNase A directly in D_2O, as described in the first experiment of the previous paragraph, were selected.

Of the 36 backbone amides that were found to be resistant to exchange in the native state, 26 of them were also found to be resistant to exchange during the desalting/concentrating steps. Of the 26 backbone amides, two exchanged too fast under the conditions employed for the labeling pulse (i.e. <1 ms). Furthermore, the protection factor on three backbone amides could not be determined well and are not included in the current study. Therefore, 21 backbone amides were used to monitor structure formation during the refolding of U_{vf}.

Fitting the Data. Simulations of the kinetic model representing the refolding of U_{vf} coupled with HD exchange were carried out using the Runge−Kutta method for numerical integration of ordinary differential equations (Press et al., 1991). The simulations were run on an IBM SP2 supercomputer (Cornell Theory Center).

RESULTS AND DISCUSSION

Refolding Pathway of U_{vf}. The refolding pathway of U_{vf} has been investigated extensively in our laboratory using the double-jump stopped flow technique (Houry et al., 1994, 1995, 1996; Houry & Scheraga, 1996). In this technique, the native (folded) protein is first unfolded for a delay time long enough to form >99% U_{vf}, but short enough so as not to form the other unfolded species. U_{vf} is then refolded, and the refolding process is monitored by different methods, including absorbance, fluorescence, circular dichroism, inhibitor binding, and, in the current study, HD exchange.

When the refolding of U_{vf} was monitored by absorbance under a wide range of GdnHCl concentrations and pHs (Houry et al., 1995), unexpected kinetics were observed. These observations indicated the presence of a hydrophobically collapsed intermediate (I_{Φ}) which has properties similar to equilibrium molten globules. In order to obtain initial information about the structure present in I_{Φ}, the refolding of U_{vf} was monitored by circular dichroism at 222 and 275 nm (Houry et al., 1996). The CD at 222 nm monitors mainly secondary structure formation, while that at 275 nm monitors the formation of tertiary contacts. The results indicated that I_{Φ} has 40−50% of the native secondary and tertiary structures. In addition, the presence of a largely unfolded intermediate (I_U) was also detected. I_U seems to differ from U_{vf} only by some local structural rearrangement.

On the basis of these observations, the refolding pathway of U_{vf} under a wide range of pHs and GdnHCl concentrations in which the native protein is baseline-folded can be represented as follows

$$U_{vf} \overset{K_i}{\rightleftharpoons} I_U \overset{K}{\rightleftharpoons} I_{\Phi} \overset{k}{\rightarrow} N \qquad (1)$$

U_{vf} first undergoes a rapid equilibration to form the largely unfolded intermediate I_U. Subsequently, I_U undergoes hydrophobic collapse to form the molten-globule-like intermediate I_{Φ}. I_{Φ} then proceeds to fold to the native state N after overcoming the rate-limiting transition state. The equilibrium constants are defined as follows: $K_i = [I_U]/[U_{vf}]$, and $K = [I_{\Phi}]/[I_U]$; k is the rate constant at which I_{Φ} proceeds to the

native state N. The apparent rate constant (k_{vf}) for the formation of N from the other species can be written as follows

$$k_{vf} = \frac{kKK_i}{1 + K_i + KK_i} \quad (2)$$

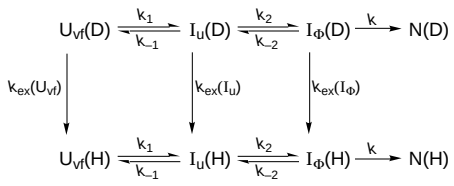
We have previously argued that $K_i \gg 1$ under all refolding conditions employed (Houry et al., 1996). Consequently, the above expression can be simplified to

$$k_{vf} = \frac{kK}{1 + K} \quad (3)$$

The equilibrium constant (K) and the rate constant (k) under a wide range of pHs and GdnHCl concentrations are given by Houry et al. (1995).

Investigating the Refolding Pathway of U_{vf} by HD Exchange. In the current study, the refolding pathway of U_{vf}, eq 1, was investigated by HD exchange in order to determine the structure present in I_U and I_Φ. All experiments were carried out at 5 °C. The native fully deuterated folded protein was first unfolded at 4.2 M GdnHCl and pH 2.0 for 1 s which allows for the formation of >99% U_{vf} without significant formation of the other unfolded species. Then U_{vf} was refolded for 6 ms, which is the shortest refolding time that could be achieved on the instrument used. Three different refolding conditions were employed: (A) 0.7 M GdnHCl and pH 3.0, (B) 2.5 M GdnHCl and pH 4.0, or (C) 4.2 M GdnHCl and pH 2.0 (where the protein remains unfolded). No exchange takes place to any significant extent during unfolding or refolding because the pH is kept low during these steps. After 6 ms, exchange (labeling) of the amide deuteriums for protons is initiated at pH 9.0 under the same GdnHCl concentration as that used for the refolding step. The labeling pulse is terminated after 20 ms by lowering the pH of the solution to 3.0 and allowing the protein to fold to the native state. The sample from C serves as a reference for the other two samples. It gives the exchange behavior of the backbone amides in U_{vf} under the conditions of our experiments. In the discussion that follows, we shall call conditions A−C the 0.7 M, 2.5 M, and 4.2 M condition, respectively.

Kinetic Models for Folding and Exchange during the Labeling Pulse. When the labeling pulse is initiated, the protein continues to fold to the native state during that pulse at the 0.7 M condition and the 2.5 M condition. Consequently, both folding and exchange are taking place simultaneously during that step under these two conditions. Hence, the kinetic model that should be considered during the labeling pulse can be written as follows



(4)

The upper row of the kinetic model represents the refolding of the deuterated protein [indicated by (D)], while the lower row represents the refolding of the protonated protein [indicated by (H)]. The rate constants k_1, k_{-1}, k_2, and k_{-2} are the fast equilibration rate constants for the formation of I_U and I_Φ. They have to satisfy the equilibrium constants given in eq 1

$$K_i = \frac{k_1}{k_{-1}} \quad (5)$$

and

$$K = \frac{k_2}{k_{-2}} \quad (6)$$

The rate constant k is the same as that given in eq 1; it describes the rate at which I_Φ proceeds to the native species. In writing the kinetic model of eq 4, we assume that the equilibration rate constants (k_1, k_{-1}, k_2, and k_{-2}) and the rate constant (k) at which I_Φ proceeds to N are the same for both the protonated and deuterated protein. $k_{ex}(U_{vf})$, $k_{ex}(I_U)$, and $k_{ex}(I_\Phi)$ are the rate constants at which the amide NDs in the deuterated protein are exchanged for protons in U_{vf}, I_U, and I_Φ, respectively. The exchange takes place in H_2O at pH 9.0 for 20 ms. Furthermore, we assume, in eq 4, that no exchange takes place from the native protein during the short 20 ms labeling pulse at the amides under consideration. Hence, no arrow is drawn for the exchange between N(D) and N(H).

Table 1 gives the values of K, k, and k_{vf}, which are defined by eqs 1, 3, 4, and 6, under the refolding and labeling conditions employed. Table 2 gives the relative concentrations of the deuterated species [$U_{vf}(D)$, $I_U(D)$, $I_\Phi(D)$, and N(D)] after 6 ms of refolding. Since labeling is initiated

Table 1: Values of K, k, and k_{vf} for the Kinetic Model of eq 1 under the Conditions Employed in the DJ−HD Experiments at 5 °C

condition	K	k (s^{-1})	k_{vf} (s^{-1})
refolding step			
0.7 M GdnHCl and pH 3.0[a]	2.47	62.0	44.1
2.5 M GdnHCl and pH 4.0[a]	0.04	29.6	1.11
exchange (labeling) step			
0.7 M GdnHCl and pH 9.0[b]	1.31	46.6	26.4
2.5 M GdnHCl and pH 9.0[b]	0.03	21.2	0.69

[a] These values of K, k, and k_{vf} for eq 1 are taken from the data of Houry et al. (1995), assuming that $K_i \gg 1$ under all conditions employed (Houry et al., 1996). The conditions used in the DJ−HD experiments include refolding at pH 3.0 and 4.0, and exchange at pH 9.0, at different GdnHCl concentrations. [b] These values are extrapolated using eqs 4 and 5 in Houry et al. (1995) (derived from data up to pH 8).

Table 2: Relative Concentrations of the Different Deuterated Species in the Kinetic Model of eq 4 after Refolding for 6 ms at the Indicated Conditions at 5 °C[a]

refolding condition	$[U_{vf}(D)]$[b]	$[I_U(D)]$	$[I_\Phi(D)]$	[N(D)]
0.7 M GdnHCl and pH 3.0	0.0	22.1	54.6	23.3
2.5 M GdnHCl and pH 4.0	0.0	95.6	3.7	0.7

[a] In the DJ−HD experiments, the protein is first unfolded to form >99% U_{vf} and then it is refolded for 6 ms prior to initiating the exchange. The concentrations given are the relative concentrations of the different deuterated species in eq 4 at the end of the 6 ms refolding step. Since HD exchange is initiated directly after 6 ms of refolding, the concentrations given are also the concentrations of the different species at zero labeling time. The relative concentrations were calculated from K and k_{vf} of Table 1 with the assumption that $K_i \gg 1$ under all conditions employed. The concentrations of the protonated species in eq 4 are zero at zero labeling time. [b] After refolding for 6 ms, all the $U_{vf}(D)$ has disappeared into a mixture of $I_U(D)$, $I_\Phi(D)$, and N(D).

directly after refolding, the relative concentrations listed in Table 2 are also the relative concentrations at the start of exchange, i.e. at zero labeling time. The concentrations of the protonated species are zero at zero labeling time. It is clear from Table 2 that the intermediate (I_Φ) can be detected only at the 0.7 M condition and not at the 2.5 M condition.

For the 4.2 M condition, we can assume that the protein is baseline-unfolded at both the refolding and the exchange steps. Salahuddin and Tanford (1970) have shown that RNase A is baseline-unfolded (100%) at 4.2 M GdnHCl, pH 2.0, and 5 °C. Furthermore, we have obtained GdnHCl transition curves for RNase A at pH 7.0 and 5 °C (Houry et al., 1996). The data indicate that, at 4.2 M GdnHCl, pH 7.0, and 5 °C, the protein is at the upper edge of the transition region next to the unfolded baseline. Therefore, we would expect the protein to be either baseline-unfolded or at the edge of the transition region at 4.2 M GdnHCl, pH 9.0, and 5 °C. Consequently, during the labeling pulse at the 4.2 M condition, I_U and I_Φ are not populated, and the only exchange that takes place in the absence of any refolding is exchange in the deuterated unfolded species, $U_{vf}(D)$,

$$U_{vf}(D) \xrightarrow{k_{ex}(U_{vf})} U_{vf}(H) \quad (7)$$

where $k_{ex}(U_{vf})$ is the same as that given in eq 4. At the start of the labeling pulse in the 4.2 M condition, the concentration of $U_{vf}(D)$ is 100%.

Amide Peak Heights. During the labeling pulse, either both folding and exchange take place according to eq 4 or only exchange takes place according to eq 7, resulting in the presence of protonated and deuterated species at the end of that pulse. The pulse is then quenched by lowering the pH to 3.0, and the protein is allowed to fold to the native state at the low pH. A DQF-COSY spectrum is obtained for the sample, and the amide peak heights (the peak heights of the NH−C^αH cross-peaks after being normalized to the Tyr 25 C^δHH′−C^ϵHH′ cross-peak) are measured. In the NMR spectrum, only the protonated backbone amides are detected, and the concentration of any given backbone amide, [NH], reflects the sum of the concentrations of the following protonated species present at the end of the labeling pulse (after 20 ms of labeling):

$$[NH] = [U_{vf}(H)]_{(20\ ms)} + [I_U(H)]_{(20\ ms)} + [I_\Phi(H)]_{(20\ ms)} + [N(H)]_{(20\ ms)} \quad (8)$$

where the (20 ms) subscript refers to the concentration of a given species after 20 ms of exchange. For the 4.2 M condition, only $U_{vf}(H)$ is present after 20 ms, i.e. $[NH] = [U_{vf}(H)]_{(20\ ms)}$.

The amide peak heights of the 21 selected amino acid residues obtained under the 0.7 M condition and under the 2.5 M condition, after being normalized to the peak heights obtained under the 4.2 M condition, are given by open and closed circles, respectively, in Figure 3. Since peak heights are directly proportional to concentrations, we can write

$$\frac{\text{HT(NH) at the 0.7 M condition}}{\text{HT(NH) at the 4.2 M condition}} = \frac{\text{[NH] at the 0.7 M condition}}{\text{[NH] at the 4.2 M condition}} \quad (9)$$

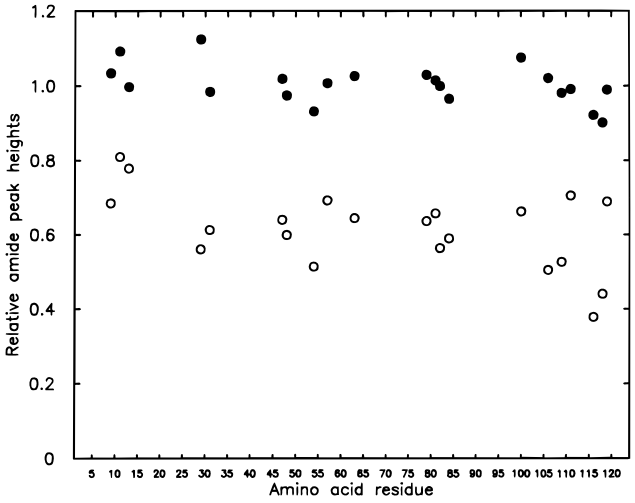


FIGURE 3: Relative amide peak heights obtained from DJ−HD experiments. The amide peak heights are the peak heights of the NH−C^αH cross-peaks referenced to the Tyr 25 C^δHH′−C^ϵHH′ cross-peak. The closed circles (●) are the amide peak heights obtained from the 2.5 M condition normalized to the amide peak heights obtained from the 4.2 M condition. The open circles (○) are the amide peak heights obtained from the 0.7 M condition normalized to the amide peak heights obtained from the 4.2 M condition. The data are averages of two repeats at each condition. The uncertainties on the amide peak heights are between 5 and 10%. The abscissa lists the sequence numbers for the amino acid residues in RNase A. There are 124 amino acid residues in the protein; 21 of these residues were used to monitor structure formation during the conformational folding of U_{vf}.

and

$$\frac{\text{HT(NH) at the 2.5 M condition}}{\text{HT(NH) at the 4.2 M condition}} = \frac{\text{[NH] at the 2.5 M condition}}{\text{[NH] at the 4.2 M condition}} \quad (10)$$

where HT(NH) refers to the peak height of a given amide obtained from the NMR spectrum and [NH] is the concentration of that amide as defined by eq 8.

Definition of Protection Factors. The extent of protection in I_U and I_Φ is calculated from the experimentally derived amide peak heights (Figure 3). A protection factor is defined (Baldwin & Roder, 1991) as the ratio of the rate constant of exchange of solvent-exposed amides measured in model compounds (k_c) to the experimentally observed exchange rate constant (k_{ex}) (i.e. protection factor = k_c/k_{ex}). If the protection factor for a given amide is close to 1, then that amide is solvent-exposed. On the other hand, if the protection factor is found to be greater than 1, then the exchange rate of that particular amide is slowed because it is either involved in some hydrogen-bonded secondary structure or buried in the interior of the protein away from the solvent.

Since U_{vf} is considered to be a completely unfolded species (Houry et al., 1994), then the exchange rate from $U_{vf}(D)$, $k_{ex}(U_{vf})$, will be similar to that obtained from model peptides after correcting for any GdnHCl effects. Hence, our protection factor will be defined as follows

$$P = \frac{k_{ex}(U_{vf})}{k_{ex}(I_x)} \quad (11)$$

with

$$k_{ex}(U_{vf}) = \beta k_c \quad (12)$$

where P is the protection factor, I_x is either I_U or I_Φ, and β is a multiplicative factor that has to be introduced in order to take into consideration the effect of GdnHCl on the intrinsic exchange rates obtained from model peptides (which were obtained in the absence of denaturant). In eq 11, both $k_{ex}(U_{vf})$ and $k_{ex}(I_x)$ must be measured under the same GdnHCl concentration.

Protection Factors in I_U. The amide peak heights obtained from the 2.5 M condition are similar to the amide peak heights obtained from the 4.2 M condition (i.e. the ratio of the amide peak heights is about 1, Figure 3). The refolding time constants at 2.5 M GdnHCl, pH 4.0 and 9.0, and 5 °C are 901 and 1449 ms, respectively (k_{vf}^{-1}, Table 1). These time constants are much larger than the length of time of the refolding and exchange pulses, 6 and 20 ms, respectively. Hence, no significant refolding takes place during the experiment at the 2.5 M condition, and consequently, about 96% (Table 2) of the exchange takes place from I_U during the labeling pulse. On the other hand, 100% of the exchange takes place from U_{vf} during the labeling pulse at the 4.2 M condition (eq 7). Hence, since the amide peak heights obtained under the two conditions (2.5 and 4.2 M) are similar, this implies that the exchange rates under the two conditions are also similar.

At this point, the effect of GdnHCl on the exchange rates must be considered. Loftus et al. (1986) have shown that the base-catalyzed exchange rates in D_2O for a poly(DL-alanine) model peptide did not change significantly between 2 and 4 M GdnHCl. Therefore, we can assume that the exchange rates at pH 9.0 in U_{vf} are the same at 2.5 and 4.2 M GdnHCl. Consequently, the protection factors, as defined by eq 11, at 2.5 M GdnHCl and pH 9.0 obtained for the observed amides in I_U are all close to 1.

Structure in I_U. Since U_{vf} is shown to be a completely unfolded species (Houry et al., 1994), this implies that, under the conditions of our experiments and within the errors of our measurements, I_U has no secondary structure which can protect from exchange the 21 backbone amides under consideration. Hence, I_U is most likely a largely unfolded species. However, it was argued before (Houry et al., 1996) that I_U might contain some local structure which gives rise to its small CD signal. This local structure was not detected by the DJ−HD experiment either because the amides involved in that structure are not part of the 21 amides under consideration or because the local structure does not provide significant protection for these backbone amides.

Protection Factors in I_Φ. The amide peak heights obtained at the 0.7 M condition are lower than those obtained at the 4.2 M condition (the ratios of the amide peak heights are less than 1, Figure 3). The amide peak heights at the 0.7 M condition normalized to those at the 4.2 M condition are typically lower than 80% due to the presence of about 20% of deuterated native protein at the start of the labeling pulse (Table 2). Furthermore, different amino acid residues have different amide peak heights, indicating different levels of protection in I_Φ. In order to obtain the protection factors in I_Φ as defined by eq 11, the exchange rate in I_Φ [$k_{ex}(I_\Phi)$] at 0.7 M GdnHCl and pH 9.0 must be determined and compared to the exchange rate in U_{vf} [$k_{ex}(U_{vf})$] at 0.7 M GdnHCl and pH 9.0.

During the labeling pulse at 0.7 M GdnHCl and pH 9.0, the refolding time constant ($k_{vf}^{-1} = 38$ ms, Table 1) and the intrinsic hydrogen−deuterium exchange time constants (k_c^{-1}) obtained from studies on model peptides (Molday et al., 1972; Bai et al., 1993) are on the same order of magnitude (on the millisecond time scale). Hence, we cannot ignore the refolding that takes place during the labeling pulse at the 0.7 M condition since both folding and exchange occur on the same time scale. Therefore, the model of eq 4 has to be used without any approximations in order to obtain an estimate of $k_{ex}(I_\Phi)$.

In order to calculate the value of $k_{ex}(I_\Phi)$ for each observed amide, the time dependence of the concentrations of each species in the kinetic model of eq 4 was simulated after making the following assumptions concerning the rate constants. The formation of I_U and I_Φ from U_{vf} was found to occur in the dead time of the various stopped flow instruments used (Houry et al., 1995, 1996). The lowest estimated dead time of these instruments is about 2 ms (Houry et al., 1994). Hence, the equilibration rate constants (k_1, k_{-1}, k_2, and k_{-2}) should have large values, at least on the order of 1000 s^{-1}, and they should satisfy the equilibrium constants as defined by eqs 5 and 6. The value of K at 0.7 M GdnHCl and pH 9.0 is given in Table 1; however, the value of K_i is not known, except that K_i must be much greater than 1 (Houry et al., 1996). For the purpose of simulating the model of eq 4, we assume that K_i is 10. The actual value used for K_i does not affect our results as long as $K_i \gg 1$. Furthermore, it was argued that I_U forms from U_{vf} before the equilibrium between I_U and I_Φ is established (Houry et al., 1996); hence, k_1 and k_{-1} should be much greater than k_2 and k_{-2}. Therefore, the following values were assigned to the equilibration rate constants: $k_1 = 10^6$ s^{-1}, $k_{-1} = 10^5$ s^{-1}, $k_2 = 1.31 \times 10^4$, and $k_{-2} = 10^4$. The actual values used for the equilibration rates do not affect the results presented below as long as the equilibration rates are larger than the exchange rates (provided that the equilibration rates satisfy the conditions discussed above). In addition, the value of k, the rate at which I_Φ proceeds to N, at 0.7 M GdnHCl and pH 9.0 is given in Table 1 (46.6 s^{-1}).

The exchange rates for the amides in I_U and U_{vf} were taken to be the same as those obtained from model peptides (k_c) after correcting for the effect of GdnHCl on the exchange rates. Loftus et al. (1986) found that the base-catalyzed exchange rates in a poly(DL-alanine) model peptide increased by a factor of about 2.3 between 0 and 2 M GdnHCl and then increased only slightly between 2 and 4 M GdnHCl. However, Loftus et al. (1986) observed the exchange of amide NHs in D_2O, while in our study, we observed the exchange of amide NDs in H_2O. Nevertheless, we will use the results of Loftus et al. (1986) as a first approximation to correct for the GdnHCl effect on the exchange rates. Therefore, $k_{ex}(U_{vf}) = k_{ex}(I_U) = 1.44k_c$ at 0.7 M GdnHCl and pH 9.0, and $k_{ex}(U_{vf}) = 2.5k_c$ at 4.2 M GdnHCl and pH 9.0. The values of 1.44 and 2.5 were obtained by linear interpolation from the data of Loftus et al. (1986). k_c at pH 9.0 was calculated according to the procedure of Bai et al. (1993) and corrected for the deuterium isotope effect on the exchange rate according to Connelly et al. (1993). It should be pointed out that the data were also treated without correcting for the effect of GdnHCl on the exchange rates, and the results obtained were similar to those presented below.

From the above analysis, all the rates (for any particular amide) given in eq 4 are now known (or approximated) except for the rate constant $k_{ex}(I_\Phi)$. To obtain the value of $k_{ex}(I_\Phi)$, this quantity was varied between 0 and $k_{ex}(U_{vf})$ using a step size of 1 s^{-1}. For each assumed value of $k_{ex}(I_\Phi)$, the rate equation for the kinetic model of eq 4 was integrated using the initial concentrations given in Table 2 to obtain the concentrations of all the species (protonated and deuterated) after 20 ms, which is the length of time of the labeling pulse. The calculated amide concentration, [NH], at the end of the 20 ms was determined according to eq 8 from the sum of the calculated concentrations of the protonated species present at the end of the 20 ms labeling pulse. This amide concentration was then divided by the amide concentration obtained from the concentration of U_{vf}-(H) after 20 ms calculated theoretically according to eq 7 for the 4.2 M condition using $k_{ex}(U_{vf}) = 2.5k_c$. The theoretically determined ratio was then compared with the experimentally obtained ratio in order to define the value of $k_{ex}(I_\Phi)$. It should be noted that the presence of 3% D_2O in the experiment during the labeling pulse does not affect the results because we are comparing ratios of peak heights. The values of $k_{ex}(I_\Phi)$ that gave theoretical ratios that were similar, within experimental error, to the experimentally determined ratios were chosen, and then the protection factors were calculated according to eq 11. The protection factors determined in this way for the amides in I_Φ are given schematically in Figures 1 and 4.

Structure in I_Φ. The protection factors for the amino acid residues in I_Φ at 0.7 M GdnHCl, pH 9.0, and 5 °C are divided into four categories as shown in Figures 1 and 4. Residues that have protection factors of less than 1.5 ($P < 1.5$) are labeled U (similar to unfolded), those with protection factors between 1.5 and 5 ($1.5 < P < 5$) L (low protection), those with protection factors between 5 and 50 ($5 < P < 50$) M (medium protection), while those with protection factors greater than 100 ($P > 100$) S (strong protection) (100 is a lower limit). The protection factors, as determined by the procedure described in the previous section, are given for the 21 selected amino acid residues (Figures 1 and 4). Figure 4 also shows the hydrogen-bonding pattern in the native state as determined from neutron and X-ray diffraction data (Wlodawer & Sjölin, 1983).

From these protection factors, a picture of the structure present in I_Φ under the conditions employed can be deduced. A protection factor which is greater than 1 for a given amide deuterium (or proton) indicates that the exchange rate of that amide is slowed in the intermediate when compared to its exchange rate in the unfolded state. This indicates that the backbone amide is either involved in hydrogen bonding as part of some secondary structure (Perrin et al., 1990) or buried in the interior of the protein away from the solvent (Englander & Kallenbach, 1984). However, it should be pointed out that it is not possible to determine whether the conformation present in the intermediate is native or nonnative (Creighton, 1991). This is one of the limitations of the pulse-labeling technique. Nevertheless, in order to have a preliminary picture of the structure present in I_Φ, we interpret the observed protection in I_Φ in terms of the secondary structure present in the native state; although, we do not (and cannot) exclude other interpretations which invoke the possible presence of nonnative conformations in the intermediate.

The structure of native RNase A has been determined by X-ray diffraction (Wlodawer et al., 1982, 1988; Wlodawer & Sjölin, 1983) and by NMR spectroscopy (Robertson et al., 1989; Rico et al., 1989, 1993; Santoro et al., 1993). The structure is given in Figure 1. To interpret the protection factors observed in I_Φ in terms of the structures (and their interactions) in the native state, the X-ray structure of Wlodawer et al. (1988) was used in order to determine the amino acid residues that are close to each other in the native molecule. A sphere with a radius of 5 Å was placed at each atom of a selected amino acid residue, and any other amino acid whose side chain or backbone was within that sphere was considered to be close in space to the selected amino acid residue.

The first helix, which consists of residues 3−13, does not seem to be stably formed in the intermediate I_Φ. Three of the residues in that helix show no protection against exchange at pH 9.0, namely Glu 9, Gln 11, and Met 13 (Figures 1 and 4). Many studies of peptide fragments representative of the amino-terminal sequence of RNase A have shown that these peptides can adopt a helical conformation in solution at low temperatures (Brown & Klee, 1971; Silverman et al., 1972; Kim & Baldwin, 1984; Osterhout et al., 1989). However, it was also shown that the stability of the helical conformation depends strongly on pH (Bierzynski et al., 1982). The helical conformation is most stable at pH 5.0, and its stability decreases at lower- and higher-pH conditions. This behavior indicates that the helical conformation of the peptide is probably stabilized by salt bridges possibly involving Glu^- 2···Arg^+ 10 and Glu^- 9···His^+ 12 (Bierzynski et al., 1982; Rico et al., 1983; Shoemaker et al., 1985; Fairman et al., 1990). Therefore, since the N-terminal helix does not seem to be stable at pH 9.0 as an independent fragment, it appears that there are no long-range interactions that would stabilize this helix in the intermediate I_Φ.

The second helix, which consists of residues 24−34, seems to be stably formed in I_Φ. Two of the residues that were detected, Met 29 and Lys 31, show strong and medium protection in the intermediate, respectively. In the native structure, the second helix is packed against the sheet regions consisting of residues 41−48, 82−84, and 95−99. Although there are no amides that could be used as probes for protection in the 95−99 β-sheet region, the other two regions (41−48 and 82−84) show low and strong levels of protection in the intermediate, possibly indicating that the second helix is stabilized by local as well as long-range interactions with these β-sheet segments.

The third helix, consisting of residues 50−60, is not formed in I_Φ as indicated by the lack of protection against exchange for Val 54 and Val 57. In addition, Val 63, which comes after the third helix and which is part of the loop region spanning the 65−72 disulfide bond, also does not exhibit any protection against exchange in the intermediate. Hence, the protein fragment between residues 50 and 72 seems to be unstructured in the intermediate. In the native state, the third helix is closely packed against the β-sheet region consisting of residues 73−79 and is also closely packed against the β-turn region consisting of residues 106−118. Met 79 shows a medium level of protection, and many of the residues in the 106−118 region also show medium and low levels of protection. This indicates that, although the regions close to the 50−72 region are stably formed in the intermediate, the local interactions are not yet established

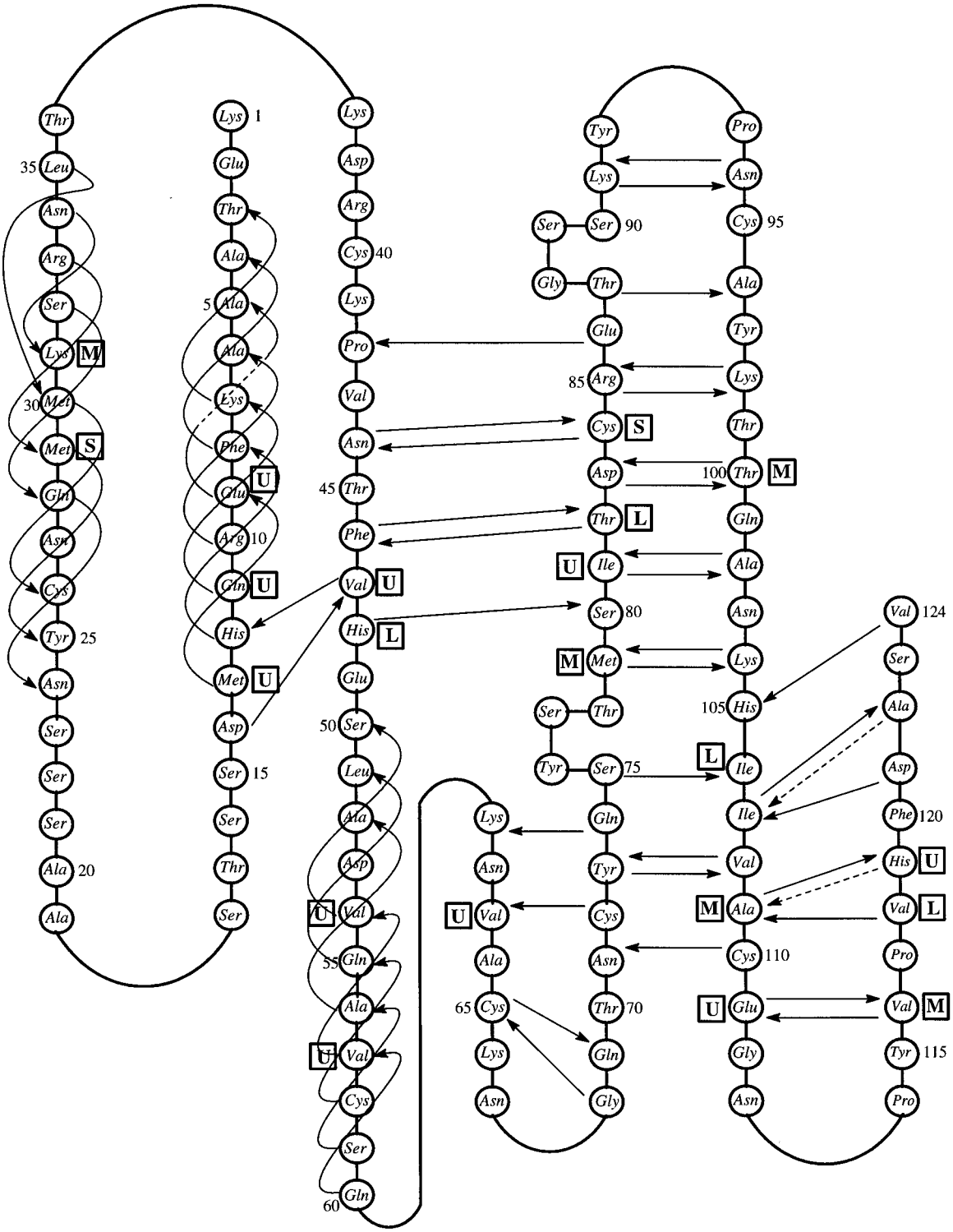


FIGURE 4: Schematic representation of RNase A showing the hydrogen-bonding pattern in the native state as obtained from neutron and X-ray diffraction data. Bonds that are shorter than 3.15 Å are drawn as solid arrows, while those between 3.15 and 3.35 Å and with the expected angles are drawn as dashed arrows. The four disulfides in the protein are not shown. This figure was modified from Figure 6 of Wlodawer and Sjölin (1983). Amino acid residues labeled U have protection factors of less than 1.5 in I_Φ ($P < 1.5$), those labeled L protection factors between 1.5 and 5 in I_Φ ($1.5 < P < 5$), those labeled M protection factors between 5 and 50 in I_Φ ($5 < P < 50$), while those labeled S protection factors greater than 100 in I_Φ ($P > 100$).

within the 50−72 region in I_Φ. Consequently, the three regions (50−72, 73−79, and 106−118) are probably not as closely packed in I_Φ as they are in the native state.

Residues 41−48 are part of a small β-sheet region in the native state. Val 47 shows no protection against exchange in I_Φ, while His 48 has a low level of protection. In the native state, Val 47 is hydrogen-bonded to residues in the first helix, and hence, the absence of protection indicates that these interactions are not present in I_Φ. This further supports our previous conclusion that the N-terminal helix is not stably formed in the intermediate. On the other hand, in the native state, His 48 is hydrogen-bonded to Ser 80 which is part of a β-sheet region that seems to be stably formed in I_Φ.

The β-sheet region extending from Met 79 to Val 118 includes many amino acid residues with low, medium, and

strong protection factors. In the native state, the β-sheet region between residues 79 and 118 contains two type VI β-turns at Tyr 92−Pro 93−Asn 94 and Asn 113− Pro 114−Tyr 115 with cis X−Pro peptide bonds. Hence, the formation of these turns in the intermediate is facilitated by the fact that the cis X−Pro peptide bonds are already present in U_{vf} (which has all the X-Pro peptide bonds in native conformations). In the native state, residues 82−84 have tertiary contacts with the second helix which is stably formed in the intermediate and has strong protection factors. Hence, both the second helix and this part of the sheet region seem to be highly structured in I_Φ. In addition, Matheson and Scheraga (1978) have postulated that the protein fragment extending from residue 106 to 118 is a chain-folding initiation site for the folding of RNase A because it is the most hydrophobic region in the protein. Since it was shown that the intermediate I_Φ is formed mainly due to hydrophobic interactions (Houry et al., 1995), then we would expect the 106−118 region to be protected in the intermediate. This is borne out by the experiment. However, the level of protection in that region is only low or medium, indicating that the structure of that region in I_Φ is probably of a dynamic nature. This implies that tertiary contacts are important for further stabilizing that region of the protein. In the native state, the 106−118 region is closely packed against the third helix (residues 50−60) and against the β-sheet region extending from residue 71 to 81. The third helix is not formed in the intermediate (as discussed above), and the 71−81 sheet region seems to be only partially formed in the intermediate. Consequently, the hydrophobically collpased 106−118 region is not further stabilized by tertiary interactions in I_Φ.

Finally, His 119 which is part of the C-terminal fragment extending from residue 119 to 124 of RNase A is not protected against exchange in I_Φ. This might indicate that the whole C-terminal fragment is not structured in the intermediate. In the native state of the protein, the C-terminal fragment is closely packed against the N-terminal helix (first helix) which is also not structured in I_Φ. Therefore, hydrophobic collapse might promote structure formation in the interior regions of the protein and might help to bring the N- and the C-termini close together; however, the structure in these termini seems to form at later stages in the folding pathway of U_{vf}. It is interesting to note that, in the region after residue 106, only the polar residues Glu 111 and His 119 are unprotected while the nonpolar residues are protected. This is consistent with hydrophobic collapse being the driving force for the formation of partial structure in that region.

By circular dichroism, we estimated that I_Φ contains 40−50% of the secondary structure and tertiary contacts present in the native state. The DJ−HD experiments detect only structures that are stable against exchange during the labeling pulse; however, circular dichroism detects structures with a wide range of stabilities whether or not these structures are able to provide protection against exchange. Hence, the structure detected by circular dichroism is an "overestimate" of the structure actually detected by DJ−HD experiments.

In summary, examination of the structure of RNaseA illustrated in Figure 1 shows that the far right side of the protein molecule consisting of the second helix and the 82−84 sheet region (and probably other sheet regions nearby) is well-structured in I_Φ. On the other hand, the left side of the protein, consisting of the N-terminal helix, the C-terminal sheet fragment, the third helix, and the 65−72 loop region, is unstructured in I_Φ. Only residues 106−118, which form the most hydrophobic segment of RNase A, adopt a flexible structure in I_Φ on the left side of the protein.

Comparison between the Structure in I_Φ and I_1. Two different intermediates have been detected on the folding pathway of the major slow-folding unfolded species U_s^{II}. Under strongly favorable folding conditions, U_s^{II} folds to the native state on the seconds time scale. On the folding pathway of U_s^{II}, an early hydrogen-bonded intermediate, labeled I_1, is initially populated (Schmid & Baldwin, 1979; Kim & Baldwin, 1980; Schmid & Blaschek, 1984; Brems & Baldwin, 1985; Udgaonkar & Baldwin, 1988, 1990, 1995). I_1 then folds to form a native-like intermediate, labeled I_N-(sII) (or just I_N) (Cook et al., 1979; Schmid, 1981, 1983, 1986; Schmid & Blaschek, 1981; Mui et al., 1985). I_N(sII) subsequently folds to the native state. The kinetic model for the refolding of U_s^{II} is

$$U_s^{II} \rightarrow I_1 \rightarrow I_N(sII) \rightarrow N \qquad (13)$$

The formation of I_1 occurs in less than 100 ms (Udgaonkar & Baldwin, 1995), compared to less than 2 ms for I_Φ. Udgaonkar and Baldwin (1988, 1990, 1995) have investigated the structure formed in the intermediate I_1 using the pulse-labeling technique. Their labeling pulse was carried out for 37 ms at 10 °C and pH 9−10 in the presence of sulfate which helps to populate the intermediate I_1. Their labeling pulse is a stronger pulse than the one used in the current study. Nevertheless, a qualitative comparison can be made between I_1 and I_Φ. In general, I_1 has a much more stable structure than I_Φ; the protection factors in I_1 are at least 1 order of magnitude larger than those in I_Φ (even under the stronger pulse conditions). The hydrogen bond network is well-established in I_1 which is not the case for I_Φ. However, there are some striking similarities between the two species. The first helix (residues 3−13) is not stably formed in both I_1 and I_Φ, while the second helix (residues 24−34) and the β-sheet region extending from residue 79 to 118 are well-structured in both species, but with much stronger protection in I_1 than in I_Φ. The most hydrophobic region of the protein consisting of residues 106−118 is strongly protected in I_1, while that region shows only medium and low protections in I_Φ. This indicates that this region of the protein is most probably hydrophobically collapsed in I_Φ with no regular structure formed, while it has a highly ordered secondary structure in I_1. Finally, the third helix (residues 50−60) is not formed in I_Φ, while it is well-structured in I_1.

The structure present in I_Φ is more flexible and dynamic than the structure present in I_1 which probably has some fixed tertiary contacts. Consequently, I_1 cannot be defined as a molten globule [as also concluded by Udgaonkar and Baldwin (1995)], since molten globules are characterized by a high degree of fluctuation compared to that of the native state. Therefore, I_Φ is more typical of molten globules than I_1.

The differences in the early-forming intermediates on the folding pathway of U_s^{II} and U_{vf} indicate that the unfolded states dictate the folding pathway that the protein has to follow to reach the native state. The folding of U_s^{II} is much slower than that of U_{vf} because of the presence of nonnative

X−Pro peptide bonds in U_s^{II}, and hence, the formation of I_1 occurs at a slower rate than the formation of I_Φ. As a result, I_1 has more time to form highly stable secondary structures and tertiary contacts. This is not the case for I_Φ which shows little protection compared to I_1. Therefore, a slower refolding rate allows for a higher cooperativity between the different structural elements of the protein, resulting in the formation of more stable (native-like) intermediates during the folding process.

CONCLUSION

By carrying out hydrogen−deuterium exchange experiments on the conformational folding pathway of U_{vf}, we obtained a preliminary picture of the structure present in the intermediates, I_U and I_Φ, which are populated on the pathway to the native state. I_U is shown to be a largely unfolded intermediate which does not contain any structure that is capable of protecting the backbone amides from exchange. On the other hand, I_Φ has well-established secondary structure involving the second helix (residues 24−34) and a large part of the β-sheet region, while the other parts of the protein remain unstructured in this intermediate. However, the levels of protection in I_Φ are generally low, indicating that the structure of this intermediate is of a dynamic nature which is typical of molten globules. The regular structure formed in I_Φ is much less than that observed in a hydrogen-bonded intermediate (I_1) populated early on the major slow-refolding pathway of the protein. In addition, the structure in I_Φ has much lower stability than that in I_1. Therefore, a slower refolding rate results in the formation of more stable intermediates.

ACKNOWLEDGMENT

We gratefully acknowledge the continuous help and invaluable advice of D. M. Rothwarf. We thank V. G. Davenport for technical assistance, C. Lester and S. Shimotakahara for advice on the NMR experiments, M. Karpinski for purifying the protein, and J.-Y. Trosset for help in identifying the closely packed amino acid residues from the X-ray structure of the protein.

REFERENCES

Anfinsen, C. B. (1973) *Science 181*, 223−230.
Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
Baldwin, R. L., & Roder, H. (1991) *Curr. Biol. 1*, 218−220.
Berger, R. L., Balko, B., & Chapman, H. F. (1968) *Rev. Sci. Instrum. 39*, 493−498.
Bierzynski, A., Kim, P. S., & Baldwin, R. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2470−2474.
Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
Brems, D. N., & Baldwin, R. L. (1985) *Biochemistry 24*, 1689−1693.
Briggs, M. S., & Roder, H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2017−2021.
Brown, J. E., & Klee, W. A. (1971) *Biochemistry 10*, 470−476.
Bycroft, M., Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488−490.
Connelly, G. P., Bai, Y., Jeng, M.-F., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 87−92.
Cook, K. H., Schmid, F. X., & Baldwin, R. L. (1979) *Proc. Natl. Acad. Sci. U. S. A. 76*, 6157−6161.
Costantino, H. R., Griebenow, K., Mishra, P., Langer, R., & Klibanov, A. M. (1995) *Biochim. Biophys. Acta 1253*, 69−74.
Creighton, T. E. (1991) *Curr. Biol. 1*, 8−10.
Desai, U. R., Osterhout, J. J., & Klibanov, A. M. (1994) *J. Am. Chem. Soc. 116*, 9420−9422.
Dodge, R. W., & Scheraga, H. A. (1996) *Biochemistry 35*, 1548−1559.
Dodge, R. W., Laity, J. H., Rothwarf, D. M., Shimotakahara, S., & Scheraga, H. A. (1994) *J. Protein Chem. 13*, 409−421.
Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521−655.
Fairman, R., Shoemaker, K. R., York, E. J., Stewart, J. M., & Baldwin, R. L. (1990) *Biophys. Chem. 37*, 107−119.
Griebenow, K., & Klibanov, A. M. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10969−10976.
Houry, W. A., & Scheraga, H. A. (1996) *Biochemistry 35*, 11719−11733.
Houry, W. A., Rothwarf, D. M., & Scheraga, H. A. (1994) *Biochemistry 33*, 2516−2530.
Houry, W. A., Rothwarf, D. M., & Scheraga, H. A. (1995) *Nat. Struct. Biol. 2*, 495−503.
Houry, W. A., Rothwarf, D. M., & Scheraga, H. A. (1996) *Biochemistry 35*, 10125−10133.
Jacobs, M. D., & Fox, R. O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 449−453.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
Jennings, P. A., & Wright, P. E. (1993) *Science 262*, 892−896.
Jones, B. E., & Matthews, C. R. (1995) *Protein Sci. 4*, 167−177.
Kim, P. S., & Baldwin, R. L. (1980) *Biochemistry 19*, 6124−6129.
Kim, P. S., & Baldwin, R. L. (1984) *Nature 307*, 329−334.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Levinthal, C. (1969) in *Mössbaur Spectroscopy in Biological Systems. Proceedings of a meeting held at Allerton House, Monticello, IL* (Debrunner, P., Tsibris, J.-M., & Munck, E., Eds.) pp 22−24, University of Illinois Press, Urbana, IL.
Loftus, D., Gbenle, G. O., Kim, P. S., & Baldwin, R. L. (1986) *Biochemistry 25*, 1428−1436.
Lu, J., & Dahlquist, F. W. (1992) *Biochemistry 31*, 4749−4756.
Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
Matheson, R. R., Jr., & Scheraga, H. A. (1978) *Macromolecules 11*, 819−829.
Molday, R. S., Englander, S. W., & Kallen, R. G. (1972) *Biochemistry 11*, 150−158.
Mui, P. W., Konishi, Y., & Scheraga, H. A. (1985) *Biochemistry 24*, 4481−4489.
Mullins, L. S., Pace, C. N., & Raushel, F. M. (1993) *Biochemistry 32*, 6152−6156.
Nozaki, Y. (1972) *Methods Enzymol. 26*, 43−50.
Osterhout, J. J., Jr., Baldwin, R. L., York, E. J., Stewart, J. M., Dyson, H. J., & Wright, P. E. (1989) *Biochemistry 28*, 7059−7064.
Perrin, C. L., Dwyer, T. J., Rebek, J., Jr., & Duff, R. J. (1990) *J. Am. Chem. Soc. 112*, 3122−3125.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800−6801.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1991) *Numerical Recipes in C*, pp 569−573, Cambridge University Press, Cambridge, U.K.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302−307.
Rico, M., Nieto, J. L., Santoro, J., Bermejo, F. J., Herranz, J., & Gallego, E. (1983) *FEBS Lett. 162*, 314−319.
Rico, M., Bruix, M., Santoro, J., Gonzalez, C., Neira, J. L., Nieto, J. L., & Herranz, J. (1989) *Eur. J. Biochem. 183*, 623−638.
Rico, M., Santoro, J., Gonzalez, C., Bruix, M., Neira, J. L., & Nieto, J. L. (1993) *Appl. Magn. Reson. 4*, 385−415.
Robertson, A. D., Purisima, E. O., Eastman, M. A., & Scheraga, H. A. (1989) *Biochemistry 28*, 5930−5938.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700−704.
Rothwarf, D. M., & Scheraga, H. A. (1993) *Biochemistry 32*, 2671−2679.
Salahuddin, A., & Tanford, C. (1970) *Biochemistry 9*, 1342−1347.
Santoro, J., González, C., Bruix, M., Neira, J. L., Nieto, J. L., Herranz, J., & Rico, M. (1993) *J. Mol. Biol. 229*, 722−734.

Schmid, F. X. (1981) *Eur. J. Biochem. 114*, 105−109.
Schmid, F. X. (1983) *Biochemistry 22*, 4690−4696.
Schmid, F. X. (1986) *FEBS Lett. 198*, 217−220.
Schmid, F. X., & Baldwin, R. L. (1979) *J. Mol. Biol. 135*, 199−215.
Schmid, F. X., & Blaschek, H. (1981) *Eur. J. Biochem. 114*, 111−117.
Schmid, F. X., & Blaschek, H. (1984) *Biochemistry 23*, 2128−2133.
Schultz, D. A., & Baldwin, R. L. (1992) *Protein Sci. 1*, 910−916.
Schultz, D. A., Schmid, F. X., & Baldwin, R. L. (1992) *Protein Sci. 1*, 917−924.
Sela, M., & Anfinsen, C. B. (1957) *Biochim. Biophys. Acta 24*, 229−235.
Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., & Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349−2353.
Silverman, D. N., Kotelchuck, D., Taylor, G. T., & Scheraga, H. A. (1972) *Arch. Biochem. Biophys. 150*, 757−766.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694−699.
Udgaonkar, J. B., & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197−8201.
Udgaonkar, J. B., & Baldwin, R. L. (1995) *Biochemistry 34*, 4088−4096.
Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H., & Clore, G. M. (1993) *Science 260*, 1110−1113.
Wlodawer, A., & Sjölin, L. (1983) *Biochemistry 22*, 2720−2728.
Wlodawer, A., Bott, R., & Sjölin, L. (1982) *J. Biol. Chem. 257*, 1325−1332.
Wlodawer, A., Svensson, L. A., Sjölin, L., & Gilliland, G. L. (1988) *Biochemistry 27*, 2705−2717.

BI961085C