Nature of the Unfolded State of Ribonuclease A: Effect of Cis−Trans X−Pro Peptide Bond Isomerization†

Walid A. Houry and Harold A. Scheraga*

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

Received March 28, 1996; Revised Manuscript Received July 9, 1996⊗

ABSTRACT: The equilibrium unfolded state of disulfide-intact bovine pancreatic ribonuclease A is a heterogeneous mixture of unfolded species. Previously, four unfolded species have been detected experimentally. They are U_{vf}, U_f, U_s^{II}, and U_s^I which have refolding time constants on the millisecond, millisecond to second, second to tens of seconds, and hundreds of seconds time scales, respectively. In the current study, the refolding pathway of the protein was investigated under favorable folding conditions of 0.58 M GdnHCl, pH 5.0, and 15 °C. In addition to the above four unfolded species, the presence of a fifth unfolded species was detected. It has a refolding time constant on the order of 2 s under the conditions employed. This new unfolded species is labeled U_m, for medium-refolding species. Single-jump refolding experiments monitored by tyrosine burial and by cytidine 2′-monophosphate inhibitor binding indicate that the different unfolded species refold to the native state along independent refolding pathways. The buildup of the different unfolded species upon unfolding of the protein from the native state was monitored by absorbance using double-jump experiments. These experiments were carried out at 15 °C and consisted of an unfolding step at 4.2 M GdnHCl and pH 2.0, followed, after a variable delay time, by a refolding step at 0.58 M GdnHCl and pH 5.0. The results of these experiments support the conclusion that the different unfolded species arise from cis−trans isomerizations at the X−Pro peptide bonds of Pro 93, 114, and 117 in the unfolded state of the protein. The rates of these isomerizations were obtained for each of these three X−Pro peptide bonds at 15 °C.

The *in vitro* protein folding reaction is the process by which a protein proceeds from its unfolded state to its well-defined native state when placed in a refolding buffer. Extensive information is available about the final product, but very little information is available about the starting reactant. The native state can be well characterized using X-ray, NMR, and other methods that provide information about the three-dimensional structure in that state. On the other hand, the unfolded state is poorly understood. It is thought to be a statistical coil in rapid fluctuation among different conformations. The unfolded state is best thought of as an ensemble of unfolded species. Recently, there have been several attempts to characterize the unfolded state of proteins structurally (Neri et al., 1992; Logan et al., 1994; Arcus et al., 1995; Zhang & Forman-Kay, 1995). Residual structure, which is native-like as well as nonnative-like, has been detected in these unfolded states (Chavez & Scheraga, 1980; Swadesh et al., 1984; Dill & Shortle, 1991; Shortle, 1993, 1996; Buckler et al., 1995). Characterization of the structure present in the unfolded state of a protein might provide some insight into the effect of such structure on the refolding process. Some of these structural elements present in the unfolded state are expected to lead to rapid folding of the protein, while others are expected to slow the refolding process.

One of the causes of the heterogeneity of the unfolded state of the protein is isomerization about X−Pro peptide bonds. This was first suggested by Brandts et al. (1975) and was then verified in many different proteins containing proline. In the amino acid sequence of bovine pancreatic ribonuclease A, there are four prolines. In the native state of the disulfide-intact protein, two of these prolines have cis X−Pro peptide bonds (Tyr 92−Pro 93 and Asn 113−Pro 114), while the other two have trans X−Pro peptide bonds (Lys 41−Pro 42 and Val 116−Pro 117) (Wlodawer et al., 1988). When disulfide-intact bovine pancreatic ribonuclease A (RNase A)[1] is unfolded, these X−Pro peptide bonds undergo cis−trans isomerization, and consequently, multiple species having different conformations about these X−Pro peptide bonds are present in the unfolded state of the protein. Different conformations about these X−Pro peptide bonds result in different refolding rates for these unfolded species.

The refolding process of RNase A has been investigated intensively (Garel & Baldwin, 1973; Schmid & Baldwin, 1979a; Henkens et al., 1980; Schmid, 1981, 1983; Lin & Brandts, 1983, 1987; Mui et al., 1985; Liu & Tsou, 1987; Kiefhaber & Schmid, 1992; Houry et al., 1994; Ybe & Kahn, 1994). These studies indicate that four phases can be observed upon refolding the protein. These four phases arise from the refolding of four unfolded species termed U_{vf}, U_f, U_s^{II}, and U_s^I, for very-fast-, fast-, major slow-, and minor slow-refolding species, respectively. Refolding of U_{vf} and

† 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,* September 1, 1996.

[1] Abbreviations: RNase A, disulfide-intact bovine pancreatic ribonuclease A; GdnHCl, guanidine hydrochloride; 2′-CMP, cytidine 2′-monophosphate free acid; MES, 2-(*N*-morpholino)ethanesulfonic acid monohydrate.

S0006-2960(96)00745-3 CCC: $12.00

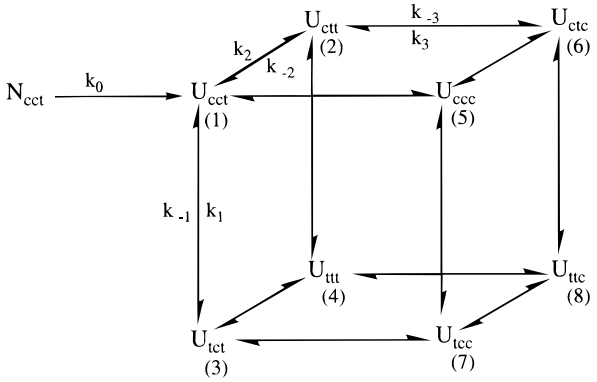

FIGURE 1: Proposed model for the unfolding of RNase A at low pH and high GdnHCl concentration (Houry et al., 1994; Dodge & Scheraga, 1996). On the basis of the model, the native state (N) unfolds conformationally to form the first unfolded species U_{cct} (i.e. U_{vf}). Then, three independent cis−trans and trans−cis isomerization events at the peptide bonds of X−Pro 93, 114, and 117 take place in the unfolded state of the protein. As a result, there are eight unfolded species which have different conformations about these three X−Pro peptide bonds. The different unfolded species are numbered sequentially, and the numbers are given in parentheses. The subscripts c and t refer to the cis or trans conformation about the X−Pro 93, 114, and 117 peptide bonds, in that order. The rate constants k_1 and k_{-1}, k_2 and k_{-2}, and k_3 and k_{-3} are the rates of isomerization of X−Pro 93, 114, and 117 peptide bonds, respectively. k_0 is the rate constant for the conformational unfolding process.

U_f takes place on the millisecond to second time scale; refolding of U_s^{II} takes place on the second to tens of seconds time scale, while refolding of U_s^{I} takes place on the hundreds of seconds time scale. U_{vf} is the unfolded species that has all the X−Pro peptide bonds in the native conformation (Houry et al., 1994, 1995), while the remaining unfolded species have one or more nonnative X−Pro peptide bonds.

Refolding studies have been carried out on several mutants of RNase A (Schultz & Baldwin, 1992; Schultz et al., 1992; Dodge et al., 1994; Dodge & Scheraga, 1996). Studies on the refolding of proline-to-alanine mutants of RNase A by Dodge and Scheraga (1996) have indicated that the X−Pro peptide bonds of Pro 93, 114, and 117 give rise to the different unfolded species. The presence of cis or trans X−Pro 93, 114, or 117 affects the refolding kinetics of the protein. On the other hand, no such effect is observed from the isomerization of the X−Pro 42 peptide bond (Dodge et al., 1994). On the basis of these results and earlier kinetic experiments (Houry et al., 1994), a model for the unfolding of RNase A at low pH and high GdnHCl concentration has been proposed (Houry et al., 1994; Dodge & Scheraga, 1996). The model is called the "Box" model (Figure 1). In this model, when the protein is unfolded, the native species undergoes a conformational unfolding event resulting in the formation of U_{vf} which has all the X−Pro peptide bonds in the native conformation. Then, from U_{vf}, cis−trans and trans−cis isomerizations take place at each of the X−Pro 93, 114, and 117 peptide bonds, resulting in the formation of the other unfolded species. These isomerization events are assumed to be independent of each other, and therefore, each unfolded species can isomerize independently to form any of three other unfolded species. This is the reason that the kinetic model is shaped like a box. The Box model fits the data of Dodge and Scheraga (1996) reasonably well.

In the current investigation, we follow the burial of the tyrosines in the protein by UV absorption to show the presence of a fifth refolding species which we call U_m, for medium-refolding species. The refolding time constant of U_m is on the order of 2 s under the conditions employed. Its presence is consistent with the Box model. In addition, the refolding kinetics of all the unfolded species were investigated by inhibitor binding. The results indicate that each unfolded species refolds along its own independent pathway to reach the native state or a native-like intermediate that can bind the inhibitor. All the phases observed upon refolding of the protein are due to parallel folding reactions.

In the current study, all experiments were carried out under conditions where the two slow-folding species, U_s^{II} and U_s^{I}, have different refolding rates and, consequently, could be monitored individually. In our previous investigation (Houry et al., 1994), the refolding conditions used were such that U_s^{II} and U_s^{I} have similar refolding rates and could not be distinguished. The formation of U_s^{II}, U_s^{I}, and the other unfolded species from the native state, upon unfolding of the protein at low pH and high GdnHCl concentration, was followed by absorbance. The results obtained further support the Box model and the proposal that X−Pro 93, 114, and 117 are responsible for the presence of the multiple unfolded species.

MATERIALS AND METHODS

Reagents. Glycine (Gly), 2-(*N*-morpholino)ethanesulfonic acid monohydrate (MES), cytidine 2′-monophosphate free acid (2′-CMP), and L-tyrosine were purchased from Sigma Chemical Co. CH_3COONa (NaAc), NaOH, and HCl were obtained from Fisher. Ultrapure guanidine hydrochloride (GdnHCl) 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-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%.

Determination of the Binding Constant of 2′-CMP to RNase A. 2′-CMP is a competitive inhibitor of the action of RNase A upon its natural substrates (e.g. ribonucleic acid). 2′-CMP was used as purchased without further purification. The extinction coefficient of 2′-CMP was calculated by obtaining wavelength scans of a solution of 88.1 μM 2′-CMP and 50 mM MES at pH 5.8 and 22 °C (room temperature) on a modified Cary model 14 spectrophotometer (Denton et al., 1982). The concentration of 2′-CMP in the solution was determined by the weight of the dry sample. From the wavelength scans, the wavelength of maximum absorption was found to be 272 nm, and the measured extinction coefficient at that wavelength was 8850 M^{-1} cm^{-1}. This is consistent with the extinction coefficients given in the literature (Beaven et al., 1955; Voet et al., 1963; Dodge et al., 1994).

Since kinetic experiments were carried out in the presence of GdnHCl (see below), we had to determine the binding constant of 2′-CMP to RNase A under similar conditions. The binding constant is defined as

$$K_b = \frac{[2'\text{-CMP}\cdot\text{RNase A}]}{[2'\text{-CMP}][\text{RNase A}]} \quad (1)$$

where [2′-CMP·RNase A] is the concentration of the complex formed between the protein and its inhibitor. K_b was determined spectrophotometrically at 0.7 M GdnHCl, pH 5.5, and 5 °C using the tandem cell technique (Anderson et al., 1968) as follows. A tandem cell (from Wilmad) having two compartments with a path length of 0.4 cm for each compartment was used. Initially, a solution of 2′-CMP containing 44 mM MES, 4.8 mM Gly, and 0.7 M GdnHCl at pH 5.5 was placed in one compartment, and a solution of RNase A containing 44 mM MES, 4.8 mM Gly, and 0.7 M GdnHCl at pH 5.5 was placed in the other compartment. An absorbance wavelength scan was obtained from 245 to 290 nm at a scan rate of 0.25 nm/s on the modified Cary model 14 spectrophotometer. Then the two solutions were mixed together in a ratio of 1:1, and the resulting solution was placed in both compartments of the tandem cell; an absorbance wavelength scan was then obtained. All experiments were carried out at 5 °C. Eleven different measurements using varying initial concentrations of 2′-CMP (0−216 μM) and of RNase A (0−141 μM) were carried out. The concentration of RNase A was determined using an extinction coefficient of 9800 M^{-1} cm^{-1} at 277.5 nm (Sela & Anfinsen, 1957), and that of 2′-CMP was determined using the extinction coefficient determined above. The wavelength of maximum difference in absorbance before and after mixing the solutions (λ_{max}) was found to be 261 nm. K_b was then calculated using the first equation given by Anderson et al. (1968). The calculated value for K_b is 4.1 $\times$ 10^4 ($\pm$0.6 $\times$ 10^4) M^{-1} at 0.7 M GdnHCl, pH 5.5, and 5 °C.

Single-Jump Refolding Kinetic Experiments. All kinetic experiments were carried out on a Hi-Tech Scientific PQ/SF-53 stopped flow instrument. The instrument and the setup associated with it have been described in detail previously (Houry et al., 1994).

Single-jump refolding experiments were carried out as follows. The unfolded protein at a concentration of 8.2 mg/mL in 4.2 M GdnHCl and 40 mM Gly at pH 1.9 was diluted with 0.22 M GdnHCl and 50 mM NaAc at pH 5.3 in a ratio of 1:10. The final refolding condition was 0.75 mg/mL RNase A, 0.58 M GdnHCl, 45 mM NaAc, and 4 mM Gly at pH 5.0. The refolding process was monitored by absorbance at a wavelength of 287 nm. Temperatures used were 5, 10, and 15 °C.

Single-jump refolding experiments were also carried out in the presence of 2′-CMP. The binding of 2′-CMP to the native species or to native-like intermediates was monitored during the refolding process. Experiments were carried out at 15 °C only. The unfolded protein at a concentration of 8.2 mg/mL in 4.2 M GdnHCl and 40 mM Gly at pH 1.9 was diluted in a ratio of 1:10 with 0.22 M GdnHCl and 50 mM NaAc at pH 5.3 containing 244 μM 2′-CMP. The final refolding buffer was 0.75 mg/mL (54 μM) RNase A, 222 μM 2′-CMP, 0.58 M GdnHCl, 45 mM NaAc, and 4 mM Gly at pH 5.0. The refolding process was monitored by absorbance at 254 nm. The change in absorbance at 254 nm arises mainly from the change in the absorbance of 2′-CMP when it binds to RNase A and not from the burial of tyrosines in the protein (Garel et al., 1976; Cook et al., 1979; Schmid & Blaschek, 1981). The burial of the tyrosines is monitored at the higher wavelength of 287 nm. When 2′-CMP binds RNase A, its absorbance decreases. At 254 nm, the formation of the complex between the inhibitor and the native state, or between the inhibitor and intermediates with native-like structures that can bind the inhibitor, is monitored and not the actual refolding process.

Single-jump kinetic experiments, whether monitored by tyrosine burial or by 2′-CMP binding, were typically repeated eight times under each condition employed. Data were collected at a rate of 0.5 ms/point for the first 10 s and then at a rate of 40 ms/point up to 600 or 900 s.

Double-Jump Refolding Kinetic Experiments. Double-jump kinetic experiments consist of two steps or two jumps. In the first jump, the protein is unfolded for a set delay time, called the unfolding time, and in the second jump, the protein is refolded. The Hi-Tech Scientific PQ/SF-53 stopped flow instrument was also used for these experiments (Houry et al., 1994).

Experiments were carried out as follows. The folded (native) protein consisting of 34.3 mg/mL RNase A in 1.5 M GdnHCl and 50 mM MES at pH 5.6 was unfolded at 4.2 M GdnHCl and pH 2.0 by 1:2.5 dilution with 5.28 M GdnHCl and 40 mM Gly at pH 1.2. The protein was allowed to unfold for a set delay time (the unfolding time), and then was refolded at 0.58 M GdnHCl and pH 5.0, and a final protein concentration of 0.9 mg/mL by diluting the unfolded protein with 0.22 M GdnHCl and 50 mM NaAc at pH 5.3 in a ratio of 1:10. The refolding process was monitored by absorbance at 287 nm. The temperature used was 15 °C. The unfolding time was varied between 0.14 and 600 s. For each unfolding time, experiments were repeated four to six times. Data collection was similar to that of the single-jump experiments.

Additional Monitoring of the Refolding of U_{vf}. The refolding process of U_{vf} was monitored at 0.7 M GdnHCl, pH 5.5, and 5 °C in the presence or absence of 2′-CMP. Experiments were carried out using the double-jump technique as follows. The folded (native) protein, 13.1 mg/mL RNase A, in 1.5 M GdnHCl and 50 mM MES at pH 5.7 was unfolded at 4.2 M GdnHCl and pH 2.0 by 1:2.5 dilution with 5.28 M GdnHCl and 40 mM Gly at pH 1.5. The protein was allowed to unfold for approximately 1 s. During this time, U_{vf} becomes populated to >99% with no significant population of the other unfolded species (Houry et al., 1995). U_{vf} was then refolded by diluting the unfolded protein, in a ratio of 1:5, with 0 M GdnHCl and 50 mM MES at pH 5.8 in the presence or absence of 216 μM 2′-CMP. The final refolding condition was 0.6 mg/mL (46 μM) RNase A, 0.7 M GdnHCl, 44 mM MES, and 5 mM Gly at pH 5.5 with or without 180 μM 2′-CMP. The refolding process was monitored by absorbance at both 254 and 287 nm. The temperature used was 5 °C. Five data sets were typically collected for each condition.

Fitting the Data. Refolding decay curves were fit using the program PLOT from New Unit (Ithaca, NY). The decay curves obtained from the refolding of U_{vf} were fit to single exponentials. Curves obtained from other single- and double-jump experiments were fit to an equation with (at most) four exponentials:

$$\Delta A(t) = \alpha_1 \exp(-t/\tau_1) + \alpha_2 \exp(-t/\tau_2) + \alpha_3 \exp(-t/\tau_3) + \alpha_4 \exp(-t/\tau_4) + \text{constant} \quad (2)$$

where $\Delta A(t)$ refers to the change in absorbance as a function of time. α_i and τ_i refer to the amplitude and time constant of phase i, respectively. The amplitudes could be either positive or negative depending on whether the decay curve is decreasing or increasing. As will be explained further and discussed below, the four observed amplitudes correspond to the refolding of the following unfolded species which are postulated to be present in the unfolded state of the protein: (1) $U_{vf} + U_f$, (2) U_m, (3) U_s^{II}, and (4) U_s^I. Hence, eq 2 can be rewritten as follows:

$$\Delta A(t) = \alpha_{(vf+f)} \exp[-t/\tau_{(vf+f)}] + \alpha_m \exp(-t/\tau_m) + \alpha_{sII} \exp(-t/\tau_{sII}) + \alpha_{sI} \exp(-t/\tau_{sI}) + \text{constant} \quad (3)$$

The reason for combining the very-fast- and fast-folding species will be presented in the Discussion. Under the refolding conditions employed, τ_{sI} was found to be on the hundreds of seconds time scale and τ_{sII} on the tens of seconds time scale, while τ_m was about 2 s and $\tau_{(vf+f)}$ on the millisecond time scale.

Since it is difficult to fit the refolding curves directly to four exponentials, the following procedure was used to fit the single-jump refolding curves. Initially, data points from 8 to 600 or 900 s were fit to a double-exponential curve $[\alpha_{sII} \exp(-t/\tau_{sII}) + \alpha_{sI} \exp(-t/\tau_{sI}) + \text{constant}_1]$ to obtain α_{sII}, τ_{sII}, α_{sI}, and τ_{sI}. Then points from 0.2 to 10 s were fit to a three-exponential curve of the form $\alpha_m \exp(-t/\tau_m) + \alpha_{sII} \exp(-t/\tau_{sII}) + \alpha_{sI} \exp(-t/\tau_{sI}) + \text{constant}_2$ while fixing α_{sII}, τ_{sII}, α_{sI}, and τ_{sI} at the previously determined values to obtain α_m and τ_m. Finally, the points from 10 ms to 0.2 or 0.5 s were fit to $\alpha_{(vf+f)} \exp[-t/\tau_{(vf+f)}] + \alpha_m \exp(-t/\tau_m) + \alpha_{sII} \exp(-t/\tau_{sII}) + \alpha_{sI} \exp(-t/\tau_{sI}) + \text{constant}_3$ while fixing α_m, τ_m, α_{sII}, τ_{sII}, α_{sI}, and τ_{sI} at the previously determined values to obtain $\alpha_{(vf+f)}$ and $\tau_{(vf+f)}$. The amplitude $\alpha_{(vf+f)}$ was corrected for the dead time of the instrument by multiplying it by $\exp[t_d/\tau_{(vf+f)}]$, where t_d is the dead time of the instrument which was found to be 2.1 ms (Houry et al., 1994).

Refolding decay curves, obtained at each unfolding time from the double-jump experiments at 15 °C, were treated by a similar procedure. In the fitting of these refolding curves, the time constants were fixed to the values obtained from single-jump refolding experiments except for the time constant $\tau_{(vf+f)}$ which was allowed to vary for reasons discussed below. The amplitude $\alpha_{(vf+f)}$ was corrected as described above.

Rate constants for the kinetic unfolding model proposed in Figure 1 were calculated from the fit to the refolding amplitudes obtained from the double-jump experiments at different unfolding times using the program ROSE developed in this laboratory (W. J. Wedemeyer and H. A. Scheraga, in preparation). The errors in the rate constants were obtained as described before (Houry et al., 1994) by using 1000 simulated data sets. The standard deviations of the refolding amplitudes included contributions from the standard deviation of several repeats, from the noise in the data, and from the error on the total amplitude.

All the errors given in this paper are at the 95% confidence limit.

RESULTS

Binding of 2′-CMP to RNase A. We examined the refolding process of RNase A in the presence of 2′-CMP in order to monitor structure formation during the folding process and, also, in order to determine whether the multiple refolding phases observed are due to parallel or sequential folding pathways. Since the refolding of RNase A was carried out in the presence of GdnHCl, the effect of GdnHCl on the binding of 2′-CMP to RNase A had to be determined in order to check whether such an experiment is feasible. To this end, the binding constant, K_b, was determined at 0.7 M GdnHCl, pH 5.5, and 5 °C (one of the conditions used for the kinetic experiments). A pH of 5.5 was chosen because the interaction between the inhibitor and the protein is greatest between pH 5.0 and 6.0 (Hummel et al., 1961).

The binding constant obtained was 4.1×10^4 ($\pm 0.6 \times 10^4$) M^{-1} at 0.7 M GdnHCl, pH 5.5, and 5 °C. This value is within the range of values reported in the literature. Anderson et al. (1968) reported a value of 18×10^4 M^{-1} at pH 5.5 and 25 °C in the absence of denaturant. Schmid and Blaschek (1981) reported a value of 50×10^4 M^{-1} at 0.2 M GdnHCl, pH 6.0, and 5 °C. Krebs et al. (1983) reported a binding constant of 4.5×10^4 M^{-1} at 0.16 M GdnHCl, pH 6.0, and 35 °C and a binding constant of 0.6×10^4 M^{-1} at 0.7 M GdnHCl, pH 6.0, and 35 °C. Dodge et al. (1994) reported a value of 2.4×10^4 M^{-1} at 0.27 M GdnHCl, pH 4.9, and 15 °C. It is clear from these data that the association between the protein and the ligand becomes stronger as the temperature is lowered and as the denaturant concentration is decreased.

Monitoring the Refolding of U_{vf}. The refolding process of U_{vf} was monitored by tyrosine burial and by 2′-CMP binding at both 287 and 254 nm using the double-jump technique (which allows the selective population of U_{vf} to >99% prior to refolding). The refolding condition employed was 0.7 M GdnHCl, pH 5.5, and 5 °C. At 287 nm, free and RNase A-bound 2′-CMP have similar absorbance properties (Hummel et al., 1961), while at 254 nm, folded and unfolded RNase A have similar absorbance properties (Schmid & Blaschek, 1981). Hence, absorbance at 287 nm monitors the burial of the tyrosines in the protein, while that at 254 nm monitors the binding of the inhibitor to the native protein with minimal contribution from tyrosine burial. When the refolding process was monitored by absorbance at 254 nm in the absence of 2′-CMP, no change was observed (Table 1), as expected. The time constants obtained when the refolding process was monitored by tyrosine burial at 287 nm in the presence or absence of 2′-CMP or at 254 nm in the presence of 2′-CMP are all similar, approximately 40 ms, within experimental error (Table 1).

When the refolding process of RNase A is monitored by 2′-CMP binding at 254 nm, it is not the formation of the folded protein that is being monitored but rather the formation of the complex between the native folded protein (or native-like intermediates) and the inhibitor. However, if the binding constant is large ($K_b \gg 1$) and if 2′-CMP is in excess of the protein, then the absorbance change at 254 nm due to 2′-CMP binding would closely reflect the refolding of RNase A [for a more detailed discussion, see Dodge et al. (1994)].

In our experiments, the refolding process of RNase A was monitored by 2′-CMP binding at 0.7 M GdnHCl, pH 5.5, and 5 °C, and, also, at 0.58 M GdnHCl, pH 5.0, and 15 °C (see below). Under these conditions, K_b is on the order of 10^4 M^{-1}. Furthermore, the concentration of 2′-CMP in the final refolding buffer was 4 times higher than that of RNase

Table 1: Time Constants for the Refolding of U_{vf} in the Presence or Absence of Inhibitor[a]

	in the presence of 2′-CMP		in the absence of 2′-CMP	
wavelength (nm)[b]	287	254	287	254
τ_{vf} (ms)	39.4 (2.9)	37.6 (2.8)[c] 38.1 (2.9)[d]	40.9 (1.2)	no change[e]

[a] Experiments were carried out at 5 °C using the double-jump technique. The folded protein in 1.5 M GdnHCl and pH 2.0 was unfolded at 4.2 M GdnHCl and pH 2.0 for 1 s, which resulted in the formation of >99% U_{vf} without significant formation of the other unfolded species. Then U_{vf} was refolded at 0.7 M GdnHCl and pH 5.5 in the presence or absence of 180 μM 2′-CMP. The final protein concentration was 46 μM (0.6 mg/mL). The refolding process was monitored by absorbance at 287 or 254 nm. All refolding curves were fit to single-exponential equations. The numbers in parentheses give the errors in the time constants at the 95% confidence limit. [b] Absorbance at 287 nm monitors mainly tyrosine burial in the protein with minimal contribution from absorbance change due to inhibitor binding. That at 254 nm monitors the change in the absorbance of the inhibitor as it binds to RNase A with minimal contribution from the absorbance change due to the burial of the tyrosines in the protein. [c] The time constant obtained from direct fitting of the absorbance change at 254 nm. The absorbance change at 254 nm monitors the formation of the complex between protein and inhibitor. [d] The time constant obtained after converting the absorbance change observed at 254 nm to the concentration of native protein according to eq 3 of Dodge et al. (1994). [e] No absorbance change (i.e. no decay curve) was observed when the refolding process was monitored by absorbance at 254 nm in the absence of inhibitor.

A. Table 1 shows that the refolding time constant obtained by fitting the absorbance change observed at 254 nm directly to a single exponential is similar to the time constant obtained after converting the absorbance change data at 254 nm to concentration of native protein according to eq 3 of Dodge et al. (1994) (37.6 vs 38.1 ms, Table 1). Hence, we are justified in assuming that monitoring the refolding of RNase A by 2′-CMP binding at 254 nm directly reflects the formation of the native protein or of native-like intermediates that can bind the inhibitor. Consequently, in the single-jump refolding experiments below, we will fit the refolding decay curves monitored by 2′-CMP binding directly, without further manipulation, to obtain the refolding amplitudes and time constants.

Single-Jump Experiments. When the refolding of RNase A is carried out under strongly favorable folding conditions (low GdnHCl and pH 5−7),[2] three kinetic phases have been observed (Cook et al., 1979; Schmid, 1983; Lin & Brandts, 1983; Mui et al., 1985). These three phases correspond to the refolding of the following unfolded species: U_f (fast-folding species), U_s^{II} (the major slow-folding species), and U_s^{I} (the minor slow-folding species). Recently, we have shown that, when single-jump refolding is carried out under unfavorable folding conditions (high GdnHCl and low pH), another phase which corresponds to the refolding of a fourth species, termed U_{vf} (for very-fast-folding), is present (Houry et al., 1994). U_{vf} folds faster than U_f. All these four species are postulated to be present in the unfolded state of the protein.

In the current study, we have carried out single-jump refolding experiments under strongly favorable folding conditions of 0.58 M GdnHCl, pH 5.0, and 5−15 °C. Under these conditions, the two slow-refolding phases, arising from the refolding of U_s^{II} and U_s^{I}, were clearly observed (Figure 2A). Furthermore, a folding phase on the millisecond time scale was also observed (Figure 2C). As will be shown below from the double-jump experiments, this phase corresponds to the refolding of both U_{vf} and U_f which have similar (but not identical) refolding rates under these conditions. The surprising result was the observation of a fourth phase which has a time constant of about 2 s (medium phase). This phase can be seen clearly in Figure 2B. A double-exponential fit to the points between 8 and 600 s does not take into account the presence of this new phase which is observed between 0.2 and 4 s.

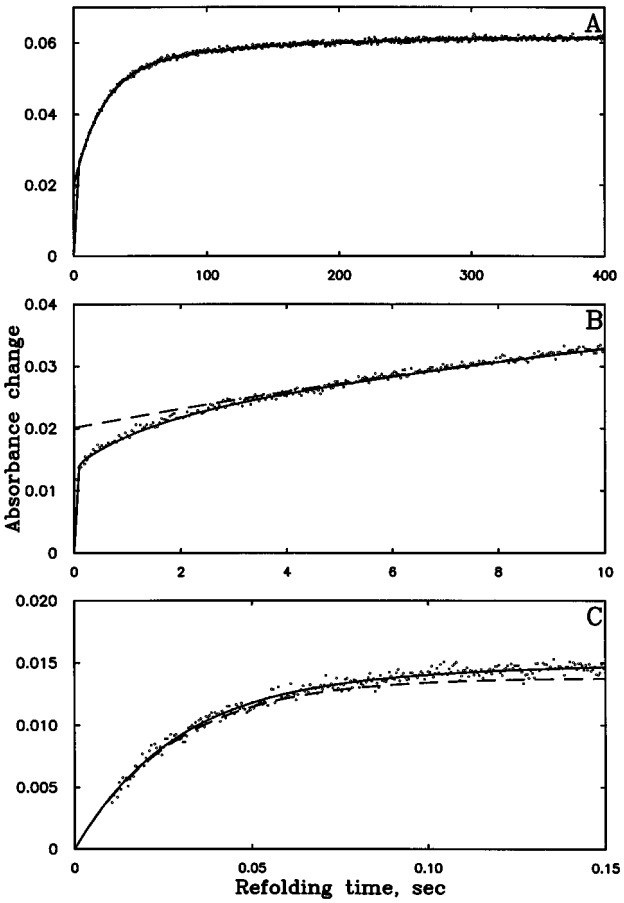

FIGURE 2: Representative curve for the refolding of RNase A under favorable folding conditions in three different time regimes (A−C). The unfolded protein in 4.2 M GdnHCl and pH 1.9 was refolded at 0.58 M GdnHCl and pH 5.0 at 15 °C. The final protein concentration was 0.75 mg/mL. In panel A, the experimental refolding decay curve is shown from 10 ms to 400 s with points plotted every 0.6 s. In panel B, the refolding curve is shown from 10 ms to 10 s with points plotted every 40 ms. In panel C, the refolding curve is shown from 10 to 150 ms with points plotted every 0.5 ms. The solid curve in all three panels is the quadruple-exponential curve $\Delta A(t) = \alpha_{(vf+f)} \exp[-t/\tau_{(vf+f)}] + \alpha_m \exp(-t/\tau_m) + \alpha_{sII} \exp(-t/\tau_{sII}) + \alpha_{sI} \exp(-t/\tau_{sI}) + \text{constant}_1$. In panels A and B, the dashed curve (− − −) is the double-exponential curve $\Delta A(t) = \alpha_{sII} \exp(-t/\tau_{sII}) + \alpha_{sI} \exp(-t/\tau_{sI}) + \text{constant}_2$. In panel C, the dashed curve is the single-exponential curve $\Delta A(t) = \alpha_{(vf+f)} \exp[-t/\tau_{(vf+f)}] + \text{constant}_3$. The average time constants and average relative amplitudes are given in Table 2. The procedure used to fit the refolding decay curve is described in Materials and Methods.

Several control experiments were carried out in order to determine whether this new phase is just a mixing artifact or a real refolding event. If the new phase were just a mixing

[2] In the text, when we refer to refolding under unfavorable folding conditions, we typically mean that the refolding process was carried out at 1.5 M GdnHCl, pH 3.0, and 15 °C. When we refer to refolding under favorable folding conditions, we typically mean that the refolding process was carried out at 0.58 M GdnHCl, pH 5.0, and 15 °C.

Table 2: Results of Single-Jump Refolding Experiments[a]

temp (°C)	Amp(vf+f)	Amp(m)	Amp(sII)	Amp(sI)	monitoring method[b]
	Relative Amplitudes				
5	27 (8)	8 (2)	47 (3)	18 (1)	TYR
10	27 (3)	9 (1)	49 (1)	15 (1)	TYR
15	21 (6)	10 (1)	53 (1)	16 (1)	TYR
15	19 (1)	8 (3)	56 (2)	17 (3)	2′-CMP
	Time Constants				
temp (°C)	$\tau_{(vf+f)}$	τ_m (s)	τ_{sII} (s)	τ_{sI} (s)	monitoring method[b]
5	46.1 (6.8)	2.0 (0.3)	29.5 (1.2)	311 (29)	TYR
10	34.3 (4.5)	1.8 (0.2)	27.1 (0.8)	188 (20)	TYR
15	27.9 (2.9)	1.6 (0.3)	22.8 (0.7)	123 (9)	TYR
15	29.6 (4.8)	1.4 (0.2)	19.9 (1.5)	121 (20)	2′-CMP
E_a (kcal/mol)[c]	8.0 (3.5)	3.6 (5.9)[d]	4.1 (1.6)	14.8 (4.0)	

[a] Single-jump refolding experiments were carried out as follows. The unfolded protein in 4.2 M GdnHCl and pH 1.9 was refolded at 0.58 M GdnHCl and pH 5.0 in the presence or absence of 222 μM 2′-CMP at the indicated temperatures. The final protein concentration was 54 μM (0.75 mg/mL). In the absence of the inhibitor, the refolding process was monitored by the change in the absorbance due to tyrosine burial at 287 nm. In the presence of inhibitor, the refolding process was monitored by the change in the absorbance of the inhibitor at 254 nm as it binds to the protein. At 254 nm, there is a minimal contribution from the absorbance change due to tyrosine burial of the protein. The amplitude and the refolding time constant of each phase are given. Subscripts vf, f, m, sII, and sI refer to the very-fast, fast, medium, major slow, and minor slow phases, respectively. Under the refolding conditions employed, the very-fast and the fast phases cannot be separated, and they are summed together in one phase labeled (vf+f) (see the Discussion). The relative amplitudes were normalized to 100. The numbers in parentheses give the errors at the 95% confidence limit. [b] The refolding process was monitored either by tyrosine burial at 287 nm in the absence of inhibitor or by inhibitor binding at 254 nm in the presence of 222 μM 2′-CMP. The two monitoring methods are referred to as TYR and 2′-CMP, respectively. [c] E_a refers to the activation energy of the refolding rate constant. [d] The activation energy for the refolding of U_m has a large uncertainty because of the large inherent errors involved in measuring τ_m (see the Discussion).

artifact, then we should be able to observe it at any wavelength as a refractive index change. No absorbance change was observed when the refolding process of RNase A was monitored at 400 nm. At this wavelength, the protein has no significant absorbance. The same refolding experiments were repeated using L-Tyrosine instead of RNase A, and when the "refolding" process was monitored by absorbance at 275 or 287 nm, no change was observed. Finally, buffer runs in the absence of protein also gave no absorbance change. On the basis of these control experiments, we can safely conclude that the new phase is an actual refolding process. The new phase has also been observed in the proline-to-alanine mutants of RNase A (Dodge & Scheraga, 1996).

Single-jump refolding experiments were also carried out in the presence of 2′-CMP. In addition to the three phases corresponding to the refolding of $U_{vf} + U_f$, U_s^{II}, and U_s^{I}, a fourth (medium) phase was also observed. The time constants and amplitudes obtained for all four phases, when the refolding of RNase A is monitored by 2′-CMP binding at 254 nm, are similar, within experimental error, to those obtained when the refolding is monitored by absorbance at 287 nm (Table 2).

Double-Jump Experiments. Double-jump experiments were carried out in order to obtain a better determination of the effect of proline isomerization on the distribution of the different unfolded species in the unfolded state of the protein. In our previous study (Houry et al., 1994), double-jump experiments were carried out under conditions in which only three distinct refolding phases were observed. In the current study, double-jump experiments were carried out under conditions in which four phases could be detected directly.

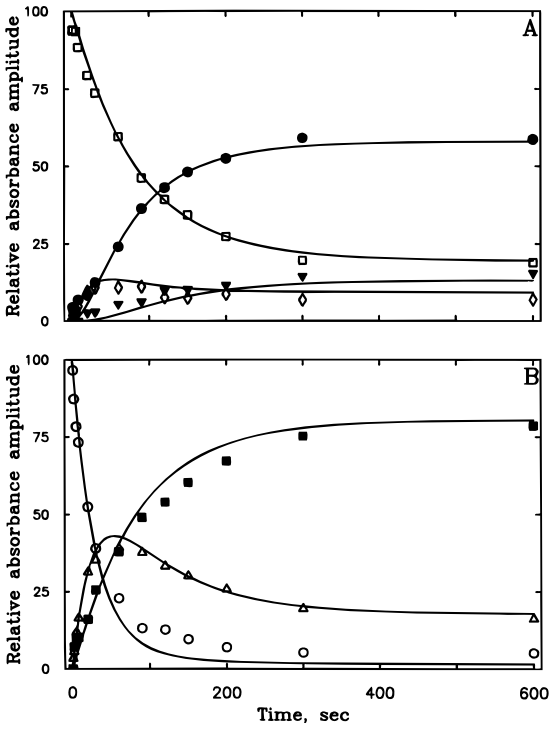

FIGURE 3: Change, with the unfolding time, of the relative absorbance amplitudes of (□) $U_{vf} + U_f$, (◇) U_m, (●) U_s^{II}, (▼) U_s^{I}, (○) U_{vf}, (△) U_f, and (■) $U_m + U_s^{II} + U_s^{I}$. Double-jump experiments were carried out as follows. The protein in 1.5 M GdnHCl and pH 5.6 was unfolded at 4.2 M GdnHCl and pH 2.0 for a set unfolding time ranging from 0.14 to 600 s, and then the protein was refolded at 0.58 M GdnHCl and pH 5.0 (A) or at 1.5 M GdnHCl and pH 3.0 (B). The final protein concentration was 0.9 mg/mL. The refolding process was monitored by absorbance at 287 nm. All experiments were carried out at 15 °C. The data of panel A are from the current investigation, while the data of panel B are from Houry et al. (1994). The solid lines are the theoretical fit to the experimental points based on the model of Figure 1 and the time constants of Table 3.

The protein was unfolded at 4.2 M GdnHCl and pH 2.0 for varying unfolding times ranging from 0.14 to 600 s and then refolded at 0.58 M GdnHCl and pH 5.0. The refolding process was monitored by absorbance at 287 nm. These studies were carried out at 15 °C. The resulting refolding decay curves were fit in order to obtain the amplitudes of each of the refolding phases at the specified unfolding time. The total absorbance amplitude, which is the sum of the amplitudes of the refolding phases, remained unchanged within an experimental error of about 9%, indicating that all the unfolded species have similar extinction coefficients. The variation of the amplitudes of each of the refolding phases as a function of unfolding time is shown in Figure 3 [the data of Figure 3A are from the current study, while the data of Figure 3B are from Houry et al. (1994)]. As the amplitude of $U_{vf} + U_f$ decreases with unfolding time, the amplitudes of U_s^{II} and U_s^{I} increase while that of the medium phase builds up and then decreases. From this behavior and from earlier results, an unfolding mechanism is proposed for RNase A.

DISCUSSION

Refolding Pathway of U_{vf}. In our previous studies (Houry et al., 1995, 1996), we have shown that, when refolding of U_{vf} is initiated, a local structural rearrangement takes place in U_{vf}, resulting in the formation of a largely unfolded intermediate (I_U). I_U then undergoes a hydrophobic collapse (a rapid pre-equilibrium) to form a molten-globule-like intermediate (I_Φ) which subsequently folds to the native state. Schematically, the folding pathway of U_{vf} can be depicted as follows:

$$U_{vf} \overset{K_i}{\rightleftharpoons} I_U \overset{K}{\rightleftharpoons} I_\Phi \overset{k}{\rightarrow} [TS] \rightarrow N \qquad (4)$$

where N is the native species and TS is the transition-state species on the refolding pathway between I_Φ and N. The equilibrium constants are defined as $K_i = [I_U]/[U_{vf}]$ and $K = [I_\Phi]/[I_U]$, while k is the rate constant for the formation of N from I_Φ. The experimentally determined rate constant (k_{vf}) for the refolding of U_{vf} is given by

$$k_{vf} = \frac{kKK_i}{1 + K_i + KK_i} \qquad (5)$$

We have previously argued that $K_i \gg 1$ under all refolding conditions employed, independent of GdnHCl concentration (Houry et al., 1996). If we assume that $K_i \gg 1$, the above expression can be simplified to

$$k_{vf} = \frac{kK}{1 + K} \qquad (6)$$

In the presence of 2′-CMP, the refolding of U_{vf} at 0.7 M GdnHCl, pH 5.5, and 5 °C was monitored by absorbance at 287 and 254 nm. A wavelength of 287 nm monitors mainly the absorbance change of the protein as the tyrosine side chains become buried in the refolding process, while a wavelength of 254 nm monitors mainly the absorbance change of the inhibitor as it binds to the protein. The coincidence of the time constants from both monitoring methods (Table 1) indicates that the active site of the protein which binds the inhibitor forms only when the tyrosines are completely buried and are present in a native or native-like environment.

Furthermore, the time constant obtained when refolding is carried out in the absence of the inhibitor monitored at 287 nm is similar to the time constants obtained in the presence of the inhibitor when monitored at 254 and 287 nm (Table 1). In addition, when the refolding of U_{vf} was monitored by 2′-CMP binding at 254 nm, no burst phase was observed within the errors of the measurements. This result is a good indication that I_Φ and TS do not bind the inhibitor to any significant extent. Hence, no native-like structure, especially in the active site region, that can bind 2′-CMP is formed in I_Φ or TS (this also applies to I_U). This agrees with our previous conclusion (Houry et al., 1995); i.e. it was found that the solvent-exposed surface area which is buried in I_Φ or TS is only about 37−46% of that of native, and consequently, the active site region which binds the inhibitor is not expected to be formed in I_Φ or TS.

Since the refolding kinetics of U_{vf} are similar in the presence or absence of 2′-CMP, then any intermediates on the refolding pathway between I_Φ and N that have native-like absorbance must have native-like structure which is responsible for binding the inhibitor. Therefore, eq 4 remains as the minimal model that satisfactorily describes the refolding process of U_{vf} under all pH, GdnHCl, and temperature conditions used.

Presence of U_m. Single-jump refolding experiments show the presence of a medium-refolding phase that refolds on a 2 s time scale (Figure 2). This phase was detected by both tyrosine burial and inhibitor binding. Furthermore, double-jump experiments show the change in the absorbance amplitude of this phase as a function of unfolding time (Figure 3A). In light of these observations, we attempt to answer the question of whether this new observed phase arises from the refolding of a fifth unfolded species which is present in the unfolded state of the protein or whether it arises from the refolding/rearrangement of an intermediate on the folding pathway of U_{vf}, U_f, U_s^{II}, or U_s^{I}.

Under strongly folding conditions, the time constants obtained for the different phases when the refolding process of RNase A is monitored by absorbance at 287 nm are similar to the time constants obtained when the refolding is monitored by 2′-CMP binding (Table 2). Furthermore, the relative amplitudes for the different kinetic phases are similar in both cases (Table 2). If U_m were an intermediate on the refolding pathway of the other known unfolded species, then three cases can be considered. If U_m, as an intermediate, either (1) bound the inhibitor as strongly as the native species or (2) did not bind the inhibitor at all, then the number of phases monitored by 2′-CMP binding would be one less than the number of phases monitored by tyrosine burial. If U_m (3) bound the inhibitor only weakly, then the relative amplitudes obtained when monitoring the refolding process of RNase A by tyrosine burial will be different from the relative amplitudes obtained when monitoring the refolding process by 2′-CMP binding, and two of the time constants will also be different. Hence, it is most likely that the medium phase arises from the refolding of a fifth unfolded species which we shall call U_m (for medium-refolding species).

Double-jump experiments lend further support to the above argument. If the new medium phase were due to the refolding/rearrangement of some intermediate on the folding pathway of one of the four known unfolded species, then the change in the absorbance amplitude of this medium phase with unfolding time would parallel the change in the amplitude of that unfolded species. In other words, α_m/α_i (where i represents any of the other species) is expected to be the same at all unfolding times (e.g. see eq A-6 or A-12 in the Appendix). From Figure 3, it is clear that this is not the case. The behavior of α_m (Figure 3A) with unfolding time is most similar to that of α_f (Figure 3B). Both build up and then decrease, but α_m/α_f does change with unfolding time (from the data of Figure 3). This also indicates that the new phase does not represent the refolding of U_f or for that matter of any of the other unfolded species. Therefore, both the single- and double-jump kinetic experiments support the conclusion that the newly observed kinetic phase arises from the refolding of a fifth unfolded species called U_m.

Refolding Rates of U_{vf} and U_f. The presence of U_{vf} can clearly be detected when the single-jump refolding experiments are carried out under unfavorable folding conditions (Houry et al., 1994). At 1.5 M GdnHCl and pH 3.0, the refolding time constant of U_{vf} is on the order of 50−40 ms between 5 and 15 °C, while that of U_f is on the order of

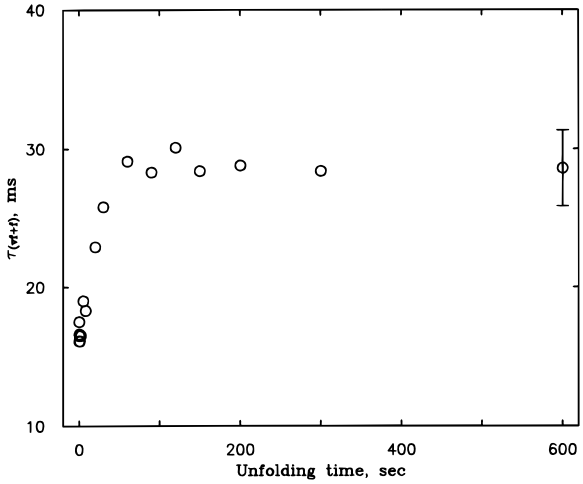

FIGURE 4: Variation of the time constant $\tau_{(vf+f)}$, obtained from double-jump experiments, plotted against unfolding time. In double-jump experiments, the folded protein in 1.5 M GdnHCl and pH 5.6 was unfolded at 4.2 M GdnHCl and pH 2.0 for varying unfolding times ranging from 0.14 to 600 s; then the protein was refolded at 0.58 M GdnHCl and pH 5.0. The final protein concentration was 0.9 mg/mL, and the temperature was kept at 15 °C. As described in Materials and Methods, when the refolding decay curves were fit at each unfolding time from double-jump experiments, the time constant $\tau_{(vf+f)}$, which corresponds to the refolding of $U_{vf} + U_f$, was not fixed in the fitting procedure. The errors in the time constants at each unfolding time were calculated at the 95% confidence limit. The average of these errors at the 95% confidence limit is shown by the error symbol on the last point in the figure.

6−4 s (Houry et al., 1994). When single-jump refolding experiments were carried out under favorable folding conditions of 0.58 M GdnHCl, pH 5.0, and 15 °C, the very-fast- and fast-refolding phases were not distinguishable. Within the quality of the experimental data, the part of the decay refolding curve between 10 ms and 0.2 s cannot be deconvoluted into two exponentials. However, when double-jump experiments were carried out, the time constant obtained from fitting that part of the refolding decay curve increased from about 17 to about 28 ms as the unfolding time increased from 0.14 to 600 s (refer to Materials and Methods for the procedure used to fit the refolding curves from double-jump experiments). This is shown in Figure 4. At early unfolding times, only U_{vf} is populated; hence, the time constant for the refolding of U_{vf} is about 17 ms at 0.58 M GdnHCl, pH 5.0, and 15 °C. However, at 600 s unfolding time, the predominant species which has a refolding time constant on the millisecond time scale is U_f. Since, at equilibrium (at infinite unfolding time), the relative ratio of $[U_{vf}]/[U_f]$ is 0.09 [from 1.6/(17.3 + 0.4), see Figure 5A], the higher refolding time constant of 28 ms reflects the time constant of U_f and not that of U_{vf}.

It is interesting to note that, although the folding of both U_{vf} and U_f seems to be mostly conformational (i.e. not rate-limited by proline isomerization), the refolding time constant of U_f has a very different dependence on the final refolding conditions than that of U_{vf}. At 0.58 M GdnHCl, pH 5.0, and 15 °C, the time constant of U_f is about 28 ms and that of U_{vf} is 17 ms. At the same pH and temperature in the presence of 1.3 M GdnHCl, the time constant of U_f is about 190 ms (Dodge & Scheraga, 1996) and that of U_{vf} is 70 ms [from the data of Houry et al. (1995)]. However, at 1.5 M GdnHCl, pH 3.0, and 15 °C, the time constant of U_f is 3.9 s and that of U_{vf} is 38 ms (Houry et al., 1994). To a first

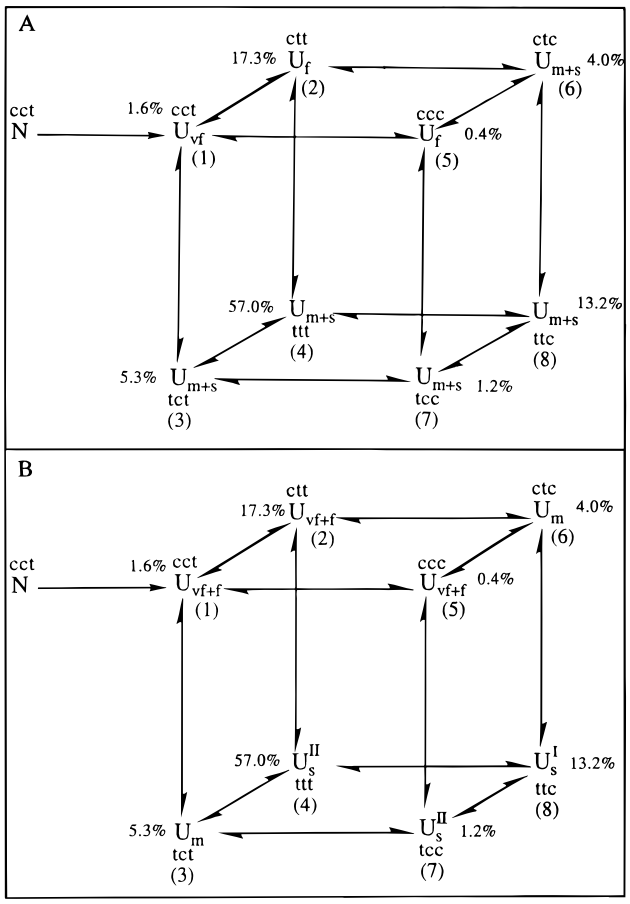

FIGURE 5: The kinetic models shown are identical to that of Figure 1. The distribution of the different unfolded species is shown when the refolding is carried out under unfavorable (A) or favorable (B) folding conditions (see eqs 7a−c and 8a−d). The species are numbered as in Figure 1. The subscripts represent the different refolding phases observed. Under unfavorable folding conditions, three phases are observed (vf, f, and m+s), while under favorable folding conditions, four phases are observed (vf+f, m, sII, and sI). c and t refer to the cis or trans conformation about X−Pro 93, 114, and 117 peptide bonds, in that order. The unfolded species are defined according to the conformations about these X−Pro peptide bonds. The percentage next to each unfolded species refers to the relative concentration of that species in the equilibrium unfolded state of the protein at 4.2 M GdnHCl, pH 2.0, and 15 °C based on the equilibrium constants of Table 3.

approximation, it seems that the refolding rates of U_f and U_{vf} have a similar GdnHCl dependence, while the refolding of U_f has a much stronger dependence on pH than that of U_{vf}. In native RNase A, several amino acid residues with carboxylic acid side chains are involved in hydrogen bonding which results in abnormally low pK_a's for these residues (Laskowski & Scheraga, 1954; Tanford & Hauenstein, 1956; Hermans & Scheraga, 1960; Antosiewicz et al., 1994; Baker & Kintanar, 1996). Therefore, the rate-limiting transition state(s) or intermediate(s) on the refolding pathway of U_f might have native-like interactions which lower the pK_a of titratable side chains significantly, resulting in a much stronger dependence of the refolding rates on pH. Such interactions leading to abnormally low pK_a's have not been observed in the intermediate (I_Φ) or in the transition state on the refolding pathway of U_{vf} (Houry et al., 1995). Further information about the nature of the rate-limiting transition state(s) or intermediate(s) on the refolding pathway of U_f could be obtained only after a systematic study of the

dependence of the refolding rate of U_f on denaturant concentration and pH.

Refolding Activation Energies. Single-jump refolding experiments under strongly folding conditions were carried out at three different temperatures (5, 10, and 15 °C) and were monitored by absorbance at 287 nm. From measurements of the refolding rate constants at each temperature, we could calculate the activation energy for the refolding of each of the unfolded species. The values are given in Table 2. In our previous study (Houry et al., 1994), we obtained the activation energy for the refolding of U_{vf} and U_f under unfavorable folding conditions. The values were 5 ± 1 and 6 ± 5 kcal/mol, respectively. These values agree within experimental error with the value obtained under favorable folding condition for $U_{vf} + U_f$ of 8 ± 4 kcal/mol (Table 2). If the refolding process were rate-limited by proline isomerization, then we would expect a large activation energy on the order of 15−20 kcal/mol (Brandts et al., 1975; Schmid & Baldwin, 1978; Houry et al., 1994). Therefore, we can conclude that the folding of both U_{vf} and U_f, under both favorable and unfavorable folding conditions, is a conformational folding process which is not rate-limited by proline isomerization.

When the refolding of RNase A was carried out under unfavorable folding conditions, only one slow phase was observed with a high refolding activation energy on the order of 16 ± 4 kcal/mol (Houry et al., 1994). However, when refolding was carried out under favorable folding conditions, two slow-folding phases were observed with very different refolding activation energies (Table 2). The refolding activation energy for U_s^{II} is 4 ± 2 kcal/mol, while that for U_s^{I} is 15 ± 4 kcal/mol. The activation energy of U_s^{I} is similar to that of the slow phase observed under unfavorable folding conditions. On the basis of the high activation energy, we expect that the refolding of U_s^{I} is rate-limited by proline isomerization. On the other hand, it is well-known that a native-like intermediate is populated on the refolding pathway of U_s^{II} [see e.g. Cook et al. 1979)]. This intermediate is also known to bind 2′-CMP but is postulated to have a nonnative (trans) Tyr 92−Pro 93 peptide bond (Cook et al., 1979; Schmid et al., 1986). Therefore, under favorable folding conditions, U_s^{II} undergoes a conformational folding process to form the native-like intermediate and then X−Pro peptide bond isomerization takes place within the intermediate to form the native state. This isomerization is silent to the monitoring methods used because both the intermediate and the native state bind 2′-CMP and both have the same absorbance properties. Hence, under favorable folding conditions, we would expect the refolding activation energy of U_s^{II} to be low, as observed (Table 2). However, when unfavorable folding conditions are employed, the native-like intermediate is not populated any more (Schmid, 1983), and the refolding of U_s^{II} becomes rate-limited by X−Pro peptide bond isomerization, resulting in a high activation energy. In this case, both U_s^{I} and U_s^{II} would have similar refolding rates and, experimentally, only one slow-folding phase would be observed with a high refolding activation energy. This is borne out by experiment. A similar conclusion about the refolding activation energy(ies) of the slow-folding species was reached by Nall et al. (1978). The limitation of the early studies of Nall et al. (1978) is that U_s^{II} and U_s^{I} were lumped together as one species (termed U_s).

The refolding rate of U_m does not show any large variation with temperature (Table 2). However, because of the difficulty in deconvoluting the exponential corresponding to the refolding of U_m from the other exponentials, large inherent errors are present in the time constants for the refolding of U_m (τ_m). Hence, we make no attempt to interpret the reason for the small variation of τ_m with temperature.

Model for the Unfolded State. On the basis of this study and our previous studies, we have been able to detect the presence of five unfolded species experimentally: U_{vf}, U_f, U_m, U_s^{II}, and U_s^{I}. Dodge and Scheraga (1996) studied the kinetics of folding of wild type and proline-to-alanine mutants of RNase A. They carried out double-jump experiments on the wild type and mutant proteins at 15 °C by unfolding at 3.9 M GdnHCl and pH 2.0 for variable times and then refolding at 1.3 M GdnHCl and pH 5.0. They were able to show that three X−Pro peptide bonds affect the refolding kinetics and contribute to the heterogeneity of the unfolded state of the protein. These three peptide bonds are X−Pro 93, 114, and 117. Furthermore, X−Pro 42 was shown not to have any effect on the refolding kinetics (Dodge et al., 1994). Using this information, Dodge and Scheraga (1996) assigned the conformation in the different unfolded species at each of the three X−Pro peptide bonds. They have shown that U_{vf} has all the X−Pro peptide bonds in the native conformation and that U_f consists of two species: one with a nonnative trans X−Pro 114 and the other with a nonnative cis X−Pro117. U_m was shown to consist of two species: one with a nonnative trans X−Pro 93 and the other with a nonnative trans X−Pro 114 and a nonnative cis X−Pro 117. U_s^{II} was shown to consist of two unfolded species, one with nonnative trans X−Pro 93 and 114, while the other has nonnative trans X−Pro 93 and nonnative cis X−Pro 117. Finally, U_s^{I} was shown to have all three X−Pro peptide bonds in a nonnative conformation.

When refolding was carried out under unfavorable folding conditions, three phases were observed: on the millisecond, second, and hundreds of seconds time scales (Houry et al., 1995). The three phases were assigned to the refolding of U_{vf}, U_f, and U_s. U_s was defined as the sum of U_s^{II} and U_s^{I}. In that study, we had not yet discovered the presence of U_m. It is most likely that, under the unfavorable folding conditions used there, U_m would fold on the same time scale as U_s since it has a nonnative X−Pro 93 or a nonnative X−Pro 114 and 117. Therefore, the U_s observed under unfavorable folding conditions is actually $U_{(m+s)} = U_m + U_s^{II} + U_s^{I}$. Figure 1 shows the Box model that was initially proposed by Houry et al. (1995) and was then verified by mutational studies of Dodge and Scheraga (1996). Under unfavorable folding conditions (1.5 M GdnHCl, pH 3.0, and 15 °C), we assign the different unfolded species as follows:

$$U_{vf} = U_{cct} = U_1 \tag{7a}$$

$$U_f = U_{ctt} + U_{ccc} = U_2 + U_5 \tag{7b}$$

$$U_{(m+s)} = U_{tct} + U_{ttt} + U_{ctc} + U_{tcc} + U_{ttc} = U_3 + U_4 + U_6 + U_7 + U_8 \tag{7c}$$

where the subscripts c or t refer to the conformation of the X−Pro peptide bond of Pro 93, 114, and 117 in that order. The numbers refer to those given in Figure 1. Under favorable folding conditions (0.58 M GdnHCl, pH 5.0, and

15 °C), we have argued above that the very-fast- and fast-refolding species cannot be deconvoluted and must be grouped together. Therefore, the four phases observed correspond to four species grouped as follows:

$$U_{(vf+f)} = U_{cct} + U_{ctt} + U_{ccc} = U_1 + U_2 + U_5 \quad (8a)$$

$$U_m = U_{tct} + U_{ctc} = U_3 + U_6 \quad (8b)$$

$$U_s^{II} = U_{ttt} + U_{tcc} = U_4 + U_7 \quad (8c)$$

$$U_s^{I} = U_{ttc} = U_8 \quad (8d)$$

Figure 5 shows the Box model and the distribution of the unfolded species when refolding is carried out under unfavorable (Figure 5A) or favorable (Figure 5B) folding conditions. The species are arranged in the same manner in Figures 1 and 5. It should be noted that the subscripts vf, f, m, sII, and sI reflect the behavior of the unfolded species upon refolding. These subscripts should be thought of as representing the different refolding phases observed rather than any given unfolded species.

Refolding Models for the Different Unfolded Species. Under unfavorable folding conditions, the refolding process of each of the unfolded species is expected to approach a two-state process since no intermediates are appreciably populated under these conditions [see Houry et al. (1994)]. Therefore, under these conditions, we can represent the folding pathway of each of the unfolded species as

$$U_i \rightarrow N \quad (9)$$

where U_i represents any of the unfolded species.

Equation 4 represents the refolding pathway of U_{vf}. Previously, we have shown that eq 4 can be used to explain the experimentally observed refolding kinetics of U_{vf} over a wide range of GdnHCl and pH conditions (Houry et al., 1995, 1996). Equation 4 reduces to eq 9 at high GdnHCl concentrations, where I_Φ is no longer populated, with $U_i = U_{vf} + I_U$. Therefore, eq 4 can be used to represent the refolding pathway of U_{vf} under all GdnHCl and pH conditions employed whether unfavorable or favorable.

Little is known about the refolding pathway of U_f or U_m. Both are proposed to contain X−Pro peptide bonds in nonnative conformations (Dodge & Scheraga, 1996). However, under favorable folding conditions of 0.58 M GdnHCl, pH 5.0, and 15 °C, U_f folds on the millisecond time scale and U_m folds in about 2 s. The refolding rates obtained when the folding process is monitored by absorbance are similar to those obtained when the folding process is monitored by 2′-CMP binding (Table 2). Hence, under favorable folding conditions, the refolding of U_f or U_m cannot be rate-limited by the slow X−Pro peptide bond isomerization. Therefore, we propose that U_f and U_m fold rapidly to native-like intermediates that have nonnative X−Pro peptide bonds, which can bind 2′-CMP and which have the same absorbance properties as that of native. These intermediates will be labeled $I_N(f)$ and $I_N(m)$, respectively.

The refolding pathway of U_s^{II} has been investigated extensively. Under favorable folding conditions, it has been shown that there are at least two intermediates which become populated on the refolding pathway of U_s^{II} (Cook et al., 1979; Schmid & Baldwin, 1979b; Kim & Baldwin, 1980; Schmid, 1983; Brems & Baldwin, 1985; Udgaonkar & Baldwin, 1988, 1990, 1995). The first intermediate that forms from U_s^{II} is a hydrogen-bonded intermediate (I_1) which protects the NHs from exchange. The second intermediate is a native-like intermediate $[I_N(sII)]$[3] which can bind inhibitor and which has absorbance properties similar to that of native. These two intermediates form sequentially from U_s^{II}.

Very little is known about the refolding pathway of U_s^{I}. The large activation energy obtained for its refolding rate under favorable folding conditions (14.8 kcal/mol, Table 2) suggests that the refolding of U_s^{I} is rate-limited by proline isomerization. Mui et al. (1985) have proposed the presence of a largely unfolded intermediate (I′) that is formed from U_s^{I} by the isomerization of a proline residue (possibly Pro 93).

On the basis of the above discussion, the independent refolding pathways for the different unfolded species, under the favorable folding condition employed here, can be represented as follows:

$$U_{vf} \rightleftharpoons I_U \rightleftharpoons I_\Phi \rightleftharpoons N \quad (10a)$$

$$U_f \rightarrow I_N(f) \rightarrow N \quad (10b)$$

$$U_m \rightarrow I_N(m) \rightarrow N \quad (10c)$$

$$U_s^{II} \rightarrow I_1 \rightarrow I_N(sII) \rightarrow N \quad (10d)$$

$$U_s^{I} \rightarrow I' \rightarrow N \quad (10e)$$

It is important to note that the above refolding models are proposed minimal ones based on the information available to date.

Relating the Refolding Amplitudes to Concentrations. In our previous study (Houry et al., 1994), we showed that the total refolding amplitude in the double-jump experiments under unfavorable folding conditions did not vary, within experimental error, with the unfolding time. This was also true in the current study where refolding was carried out under favorable folding conditions and where U_m, U_s^{II}, and U_s^{I} were individually detected. Therefore, we can reasonably assume that all the unfolded species have similar absorbances. Hence, the absorbance properties of the unfolded species, the intermediates, and the native state can be related as follows:

$$\epsilon_U = \epsilon(U_{vf}) = \epsilon(U_f) = \epsilon(U_m) = \epsilon(U_s^{II}) = \epsilon(U_s^{I}) = \epsilon(I_U) = \epsilon(I_\Phi) = \epsilon(I_1) = \epsilon(I') \quad (11)$$

$$\epsilon_N = \epsilon(N) = \epsilon[I_N(f)] = \epsilon[I_N(m)] = \epsilon[I_N(sII)] \quad (12)$$

where U is a generic species representative of any of the unfolded species and ϵ represents the extinction coefficient. ϵ_U and ϵ_N are the extinction coefficients of the unfolded and the native species, respectively.

The unfolded state of RNase A consists of all the different unfolded species given above. When refolding of this ensemble of unfolded species is monitored by absorbance, each of the unfolded species appears to fold in a single-exponential decay curve to the native or native-like state.

[3] $I_N(sII)$ has been referred to in the literature as I_N [see, e.g., Cook et al. (1979)].

Under strongly folding conditions, each of the unfolded species will fold according to eqs 10a−e. In the Appendix, we show that the change in absorbance as a function of refolding time for each of the refolding eqs 10a−e will follow an apparent single-exponential decay curve. For eqs 10a−c, this follows directly from the absorbance properties of the intermediates (eqs 11 and 12). For eqs 10d,e, in addition to using the absorbance properties of the intermediates, the following approximations were also used. For eq 10d, the refolding step from U_s^{II} to I_1 was assumed to be much faster than that from I_1 to I_N(sII). This agrees with what is known in the literature about the properties of I_1 and I_N(sII) (Schmid & Baldwin, 1979b; Kim & Baldwin, 1980; Udgaonkar & Baldwin, 1988, 1990, 1995). I_1 is an early-folding intermediate on the refolding pathway of U_s^{II}, while I_N(sII) forms on the second time scale under favorable folding conditions. For eq 10e, the conformational folding step from I′ to N was assumed to be faster than the isomerization step from U_s^I to I′. This agrees with the proposal of Mui et al. (1985) for the refolding of U_s^I at low denaturant concentrations.

The reason why the refolding process of each of the unfolded species appears to be a single-exponential decay curve can be rationalized as follows (the mathematical treatment is given in the Appendix). During the refolding of U_{vf}, the absorbance change observed is due to the formation of the native species from all the other species given in eq 10a with an apparent rate given by k_{vf} (eq 5). Since the equilibration between U_{vf}, I_U, and I_Φ occurs within the dead time of the instrument, and since all three species have the same extinction coefficient (ϵ_U), then the refolding of U_{vf} will appear to be a single-exponential process with no burst phase. For the refolding of U_f or U_m, only the reaction from U_f to I_N(f) or from U_m to I_N(m) is monitored by absorbance. The subsequent step from I_N(f) to N or from I_N(m) to N is silent, because ϵ(N) = ϵ[I_N(f)] = ϵ[I_N(m)]. For the refolding of U_s^{II}, only the reaction from I_1 to I_N(sII) is monitored by absorbance. The step from U_s^{II} to I_1 is too fast with respect to the reaction from I_1 to I_N(sII), while the step from I_N(sII) to N cannot be detected because ϵ(N) = ϵ[I_N(sII)]. Finally, for the refolding of U_s^I, the slow step is the rate-limiting step, and the apparent reaction monitored is that from U_s^I to I′. Therefore, the refolding processes of each of the unfolded species will appear to be a two-state reaction resulting in single-exponential decay curves.

The amplitude obtained from the apparent single-exponential decay curve for the refolding of each of the unfolded species is equal to the concentration of that species at zero refolding time multiplied by ($\epsilon_U - \epsilon_N$) (see the Appendix). Hence, we can write

$$\alpha_{vf}:\alpha_f:\alpha_m:\alpha_{sII}:\alpha_{sI} = [U_{vf}]_0:[U_f]_0:[U_m]_0:[U_s^{II}]_0:[U_s^I]_0 \quad (13)$$

where α is the absorbance amplitude obtained from the apparent single-exponential decay curve for the refolding of each species and $[U_i]_0$ is the concentration of species U_i at zero refolding time.

Isomerization Rates of the X−Pro Peptide Bonds in the Unfolded State of RNase A at 15 °C. On the basis of the above discussion, the refolding amplitudes obtained from the double-jump experiments upon fitting the refolding decay curves at each unfolding time are directly proportional to the relative concentrations of the different unfolded species that are present in the unfolded state at that given unfolding time. The variation of the relative absorbance amplitudes with the unfolding time is given in Figure 3. Figure 3A shows the relative amplitudes when the refolding step is carried out under favorable folding conditions of 0.58 M GdnHCl, pH 5.0, and 15 °C, while Figure 3B shows the relative amplitudes when the refolding step is carried out under unfavorable folding conditions of 1.5 M GdnHCl, pH 3.0, and 15 °C. The data for Figure 3B are taken from our previous study (Houry et al., 1994). In both studies, the unfolding step was carried out at 4.2 M GdnHCl and pH 2.0.

Under favorable folding conditions, four phases were observed corresponding to the refolding of $U_{(vf+f)}$, U_m, U_s^{II}, and U_s^I. Under unfavorable folding conditions, only three phases were observed corresponding to the refolding of U_{vf}, U_f, and $U_{(m+s)}$. In the Box model of Figure 1, there are eight unfolded species. These species were grouped according to eqs 7a−c and 8a−d, depending on the refolding conditions used. This is also shown in panels A and B of Figure 5. The amplitudes in panels A and B of Figure 3 were then fit simultaneously to the Box model of Figure 1 to obtain the cis−trans and the trans−cis isomerization rate constants at each of the three X−Pro 93, 114, and 117 peptide bonds in the unfolded state of the protein at 15 °C. The isomerization rate constants are given in Table 3, and the resulting theoretical decay curves are shown as solid lines in Figure 3. In the fitting of the data, k_0 was fixed at 28.57 s^{-1} (i.e. τ = 35.0 ms) which is the value obtained from single-jump unfolding experiments (Houry et al., 1994). All other rate constants were allowed to vary.

The theoretical curves drawn in Figure 3 fit the experimental points reasonably well, indicating that the Box model of Figure 1 is a good first approximation for the nature of the distribution of the different unfolded species in the unfolded state of the protein. The isomerizations around X−Pro 93, 114, and 117 are essential for characterizing the heterogeneity of the unfolded state. This was shown explicitly by the proline-to-alanine mutants of Dodge and Scheraga (1996). Dodge and Scheraga carried out double-jump studies on the wild type protein and on the mutants P93A, P114A, and P117A. They fit their data to the Box model of Figure 1 to obtain the time constants for the isomerizations about the X−Pro peptide bonds. Their values are also given in Table 3. From Table 3, it is clear that the isomerization time constants and the equilibrium constants of Dodge and Scheraga agree with those obtained in the current study within experimental error.

It should be pointed out that, even though the isomerization time constants for X−Pro 117 obtained in the current study are similar to those of Dodge and Scheraga (1996) within experimental error, the best fit values themselves are twice as large as those of Dodge and Scheraga. There is a large uncertainty in the isomerization time constants of X−Pro 117. Figure 6 shows the behavior of all the eight unfolded species as a function of unfolding time when the protein is unfolded at 4.2 M GdnHCl, pH 2.0, and 15 °C. The curves drawn are theoretical ones generated on the basis of the Box model of Figure 1 and the time constants of Table 3. As shown in Figure 6, the species formed by the isomerization of X−Pro 117, namely U_5, U_6, U_7, and U_8, have low amplitudes throughout the unfolding time, and hence, when being fit to the Box model, the isomerization rates of X−Pro 117 are determined mainly by experimental data that have

Table 3: Results of the Fit to the Box Model[a]

isomerization time constants at 15 °C (s)	X−Pro 93		X−Pro 114		X−Pro 117	
	c → t $(k_1)^{-1}$	t → c $(k_{-1})^{-1}$	c → t $(k_2)^{-1}$	t → c $(k_{-2})^{-1}$	c → t $(k_3)^{-1}$	t → c $(k_{-3})^{-1}$
current study	120 (15)	389 (130)	43 (8)	468 (232)	125 (79)	539 (270)
Dodge and Scheraga[b]	130	500	50	385	67	204
equilibrium constants at 15 °C	$K_1 = k_1/k_{-1}$		$K_2 = k_2/k_{-2}$		$K_3 = k_3/k_{-3}$	
current study	3.3 (1.2)		10.8 (5.7)		4.3 (3.4)	
Dodge and Scheraga[b]	3.9		7.7		3.1	

[a] The relative refolding absorbance amplitudes obtained from the double-jump experiments at different unfolding times were fit to the Box model shown in Figure 1. In the double-jump experiments, the folded protein in 1.5 M GdnHCl and pH 5.6 was unfolded at 4.2 M and pH 2.0. The protein was kept unfolded for variable amounts of time ranging from 0.14 to 600 s. The unfolded protein was then refolded at 0.58 M GdnHCl and pH 5.0 (data of Figure 3A) or at 1.5 M GdnHCl and pH 3.0 (data of Figure 3B). The refolding process was monitored by absorbance. All experiments were carried out at 15 °C. The different refolding species observed were grouped according to eqs 7a−c and 8a−d. The data of Figure 3B are from Houry et al. (1994). The rate constants refer to those shown in Figure 1. In the fitting of the data, all the rate constants were allowed to vary except for k_0 which was fixed at 28.57 s^{-1} (Houry et al., 1994). The numbers in parentheses give the error at the 95% confidence limit. [b] The values listed are those obtained by Dodge and Scheraga (1996).

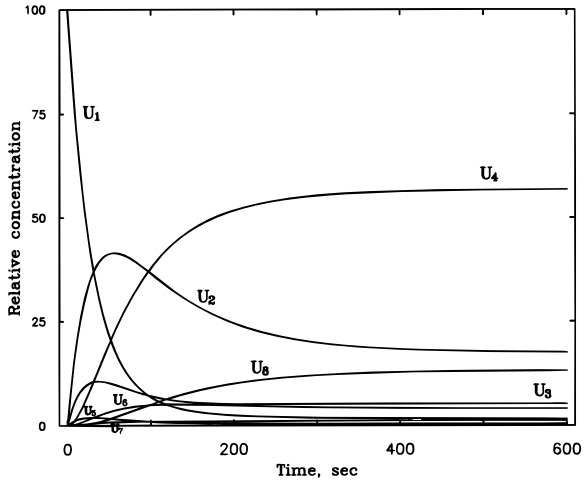

FIGURE 6: Change in the relative concentration of each of the eight unfolded species of Figure 1 with unfolding time when RNase A is unfolded at 4.2 M GdnHCl, pH 2.0, and 15 °C. The curves were generated on the basis of the kinetic model of Figure 1 and the isomerization time constants of Table 3. The unfolded species are numbered according to Figure 1.

large errors because of their low amplitudes. Therefore, the difference in the isomerization rates of X−Pro 117 between the current study and that of Dodge and Scheraga can arise solely from experimental error.

Another possible source for the discrepancy can be the assumption made about the refolding mechanism of the mutant proteins. In these mutants, replacing a proline with an alanine affects the absorbance properties of the protein and consequently can affect the absorbance properties of the intermediates present on the refolding pathways. Two of the tyrosines in the protein are next to prolines, namely Tyr 92−Pro 93 and Pro 114−Tyr 115. Mutating each of these two prolines to alanine results in an increase in the extinction coefficient of the protein [see Table 1 of Dodge and Scheraga (1996)]. Mutating the other two prolines has no substantial effect on the extinction coefficient of the protein. Therefore, intermediates present on the refolding pathway of the mutant proteins might not have absorbance properties similar to those of the native or unfolded protein. Consequently, the relationship between the refolding amplitude and concentration would have to be examined more carefully.

Finally, the Box model of Figure 1 is a minimal model that was proposed on the basis of a kinetic analysis of the data (Houry et al., 1994), and the model was then verified by mutational studies (Dodge & Scheraga, 1996). It is important to recognize that each of the unfolded species is an ensemble of unfolded species that are in rapid equilibrium with each other. The model fits well to our data of Figure 3 and to the data of Dodge and Scheraga (1996) for the wild type, P93A, and P114A mutant proteins whether using the isomerization rates of Dodge and Scheraga or the isomerization rates of the current study (results not shown). However, the Box model does not fit well to the double-jump refolding amplitudes of the P117A mutant protein irrespective of whether the isomerization rates of Dodge and Scheraga [Figure 10 of Dodge and Scheraga (1996)] or the isomerization rates of the current study (not shown) are used. This might indicate that there are other events or isomerizations that occur in the unfolded state of the protein that give rise to these multiple unfolded species. At this stage, we can only speculate as to the nature of these events. They might include, for example, disulfide bond isomerization or other X−Y peptide bond isomerizations where Y is not a proline. Further mutational studies will be needed to determine them. Consequently, the Box model would have to be expanded further.

Equilibrium Distribution of the X−Pro Peptide Bonds. Table 3 lists the equilibrium constants obtained for the cis−trans isomerization at each of the three X−Pro peptide bonds (Tyr 92−Pro 93, Asn 113−Pro 114, and Val 116−Pro 117) in the unfolded protein at 4.2 M GdnHCl, pH 2.0, and 15 °C. The values obtained in the current study are similar to those obtained by Dodge and Scheraga (1996). On the basis of the equilibrium values of the current study, X−Pro 93 has 68−82% of the peptide bonds in a trans conformation and 18−32% in a cis conformation in the unfolded state; while for X−Pro 114, 84−94% are in a trans conformation and 6−16% are in a cis conformation. Hence, the X−Pro 114 peptide bond has a lower percentage of cis isomers than X−Pro 93. Because of the large errors in the isomerization rates of X−Pro 117, no reasonable estimate can be made for the percentage of cis or trans isomers at that peptide bond.

Adler and Scheraga (1990) have found by NMR spectroscopy that, in the unfolded state of RNase A, 60% of the

X−Pro 93 peptide bonds are in the trans conformation and that 63% of the X−Pro 114 peptide bonds are in the trans conformation. The standard deviation on each value is 7%. At a 95% confidence limit, the values obtained by Adler and Scheraga (1990) would be similar to the values obtained in the current study. Nevertheless, it should be pointed out that, in their study, the protein was heat-unfolded at pH 2.0. There might be a difference in the distribution of the cis and trans conformations of the X−Pro peptide bonds between heat-unfolded and denaturant-unfolded protein. Stimson et al. (1982) obtained the NMR spectrum of the fragment corresponding to residues 105−124 of RNase A. They found that 12% of the Asn 113−Pro 114 peptide bonds in this fragment are in a cis conformation. They also found that, in the tripeptide Tyr−Pro−Asn, which corresponds to the sequence Tyr 92−Pro 93−Asn 94 in RNase A, 29% of the Tyr−Pro peptide bonds are in the cis conformation. Their findings are in agreement with the equilibrium values obtained in the current study for X−Pro 93 and 114. From studies on model peptides (Grathwohl & Wuthrich, 1976; Dyson et al., 1988; Yao et al., 1994), it was found that the presence of an aromatic residue before a proline residue results in a higher population of the cis X−Pro peptide bond conformation. This is consistent with the cis-trans distribution of Tyr 92−Pro 93 compared to that of Asn 113−Pro 114 as given above.

In summary, the values obtained for the equilibrium constants about the three X−Pro peptide bonds under consideration are in agreement with observations from studies on model peptides. Using these equilibrium constants, the relative concentrations of the different unfolded species in the equilibrium unfolded state of the protein at 4.2 M GdnHCl, pH 2.0, and 15 °C can be calculated

$$[U_1]:[U_2]:[U_3]:[U_4]:[U_5]:[U_6]:[U_7]:[U_8] = 1.6:17.3:5.3:57.0:0.4:4.0:1.2:13.2 \quad (14)$$

The relative percentages are also given in Figure 5.

Contribution of the X−Pro 42 Peptide Bond. In addition to the X−Pro peptide bonds of Pro 93, 114, and 117, the X−Pro 42 peptide bond is also expected to isomerize when the protein is unfolded. However, since the conformation around the X−Pro 42 peptide bond does not affect the refolding kinetics (Dodge et al., 1994), the isomerization around X−Pro 42 would be "silent" in the double-jump experiments. However, this does not exclude the fact that the isomerization around X−Pro 42 contributes to the heterogeneity of the unfolded state of the protein. Figure 7 shows the kinetic model for the unfolding of RNase A at both low pH and high GdnHCl concentration when four independent isomerization events take place in the unfolded state of the protein at X−Pro 42, 93, 114, and 117. This results in the formation of 16 unfolded species. Since the isomerization about the X−Pro 42 peptide bond does not affect the refolding kinetics, the eight species (U_1−U_8) (solid box) will have the same refolding behavior as the other eight species (U_9−U_{16}) (dashed box). At this stage of our investigation, we are not able to obtain the rates of isomerization about the X−Pro 42 peptide bond. A structure determination of the unfolded state of RNase A by heteronuclear NMR would provide further information about the effect of X−Pro cis−trans peptide bond isomerization on the unfolded state of the protein.

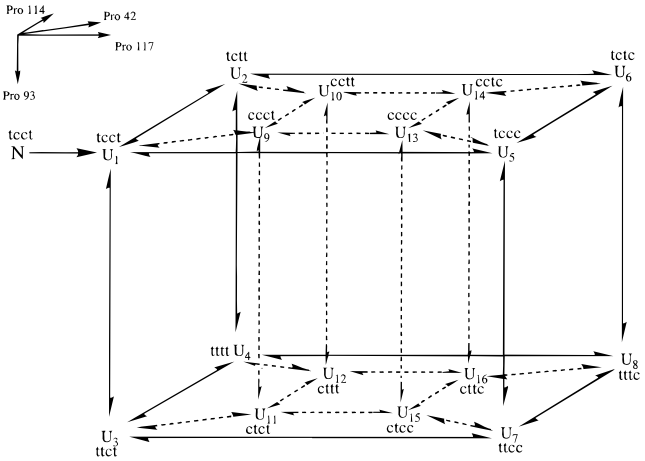

FIGURE 7: Proposed kinetic model for the unfolding of RNase A when there are four independent isomerization events taking place in the unfolded state of the protein. The protein conformationally unfolds from the native state (N) to the first unfolded species U_1 (i.e. U_{vf}); then independent isomerizations take place at X−Pro 42, 93, 114, and 117 peptide bonds. c and t refer to the cis or trans conformation of X−Pro 42, 93, 114, and 117 peptide bonds, in that order. The model consists of a box inside another box. Species U_1−U_8 (the outer solid box) and species U_9−U_{16} (the inner dashed box) are the same as the unfolded species of Figure 1 but with a specified trans or cis X−Pro 42 peptide bond, respectively. The arrows above the figure help to show the directionality of the isomerization of each of the four X−Pro peptide bonds in the model.

CONCLUSION

Through the study of the refolding process of RNase A under favorable folding conditions, a new medium-folding phase was observed in addition to the very-fast-, fast-, and slow-folding phases. The new phase arises from the refolding of a distinct unfolded species called U_m. Consequently, there are at least five experimentally observed species in the unfolded state of RNase A: U_{vf}, U_f, U_m, U_s^{II}, and U_s^{I}. The five experimentally observed unfolded species correspond to eight theoretically possible species which have different conformations at each of the X−Pro 93, 114, and 117 peptide bonds. U_{vf} has all three X−Pro peptide bonds in native conformations, while U_s^{I} has all three X−Pro peptide bonds in nonnative conformations. It is observed that the refolding rate of U_f has a much stronger pH dependence than that of U_{vf}. The results indicate that each of the unfolded species refolds to the native state along distinct folding pathways at different refolding rates.

ACKNOWLEDGMENT

We thank W. J. Wedemeyer for writing the fitting programs used in this study and D. M. Rothwarf for invaluable advice and for careful reading of the manuscript. We also thank R. W. Dodge for helpful discussions and M. Karpinski for purifying the protein.

APPENDIX

Derivations of the Relations between the Absorbance Refolding Amplitudes and Concentrations

In this Appendix, we derive the mathematical relations between the experimentally determined absorption amplitude for the refolding of each of the unfolded species and the concentration of that species. For this purpose, we discuss the refolding pathway of each unfolded species separately. In the following, the absorbance path length is assumed to be 1 cm.

U_{vf}. Experimentally, the total absorbance change monitored as a function of refolding time (t) is the sum of the absorbance changes of all the species present on the refolding pathway. This can be represented as follows

$$\text{total absorbance}(t) = \sum_{i} \text{absorbance}(t) \text{ of species i} \tag{A-1}$$

The refolding pathway of U_{vf} is given in eqs 4 and 10a. From eq A-1, we can write

$$\text{total absorbance}(t) = \text{absorbance}(U_{vf}) + \text{absorbance}(I_U) + \text{absorbance}(I_\Phi) + \text{absorbance}(N) \tag{A-2}$$

We assume that, at zero refolding time, only U_{vf} is present at a concentration of $[U_{vf}]_0$. It can easily be shown that

$$[U_{vf}] = [U_{vf}]_0[k_{vf}/(kKK_i)]e^{-k_{vf}t} \tag{A-3a}$$

$$[I_U] = [U_{vf}]_0[k_{vf}/(kK)]e^{-k_{vf}t} \tag{A-3b}$$

$$[I_\Phi] = [U_{vf}]_0(k_{vf}/k)e^{-k_{vf}t} \tag{A-3c}$$

$$[N] = [U_{vf}]_0(1 - e^{-k_{vf}t}) \tag{A-3d}$$

where k_{vf}, k, K, and K_i are defined in eqs 4 and 5. Substituting the above equations into eq A-2 and using the relations given by eqs 11 and 12, we obtain

$$\text{total absorbance}(t) = (\epsilon_U - \epsilon_N)[U_{vf}]_0e^{-k_{vf}t} + \epsilon_N [U_{vf}]_0 \tag{A-4}$$

The above expression is the experimentally observed absorbance decay curve upon refolding U_{vf}. Hence, the experimentally determined amplitude for the refolding of U_{vf} is $(\epsilon_U - \epsilon_N)[U_{vf}]_0$.

U_f, U_m, and U_s^I. The refolding pathway of these three species is given by a three-state sequential model (eqs 10b,c,e). In a general form, the model is

$$U \xrightarrow{k_1} I \xrightarrow{k_2} N \tag{A-5}$$

U represents U_f, U_m, or U_s^I, while I represents $I_N(f)$, $I_N(m)$, or I′. The solution to the above model is given by Szabo (1969). At zero refolding time, only U is present at a concentration of $[U]_0$. It can be shown that the expression for each of the species as a function of refolding time is

$$[U] = [U]_0e^{-k_1t} \tag{A-6a}$$

$$[I] = [U]_0 [k_1/(k_2 - k_1)](e^{-k_1t} - e^{-k_2t}) \tag{A-6b}$$

$$[N] = [U]_0\{(1 - e^{-k_1t}) - [k_1/(k_2 - k_1)](e^{-k_1t} - e^{-k_2t})\} \tag{A-6c}$$

The total absorbance change as a function of the refolding time for eq A-5 is given by

$$\text{total absorbance}(t) = \epsilon_U[U] + \epsilon_I[I] + \epsilon_N[N] \tag{A-7}$$

In the case of U_f and U_m, $\epsilon_I = \epsilon_N$ (eq 12). Therefore, eq A-7 becomes

$$\text{total absorbance}(t) = (\epsilon_U - \epsilon_N)[U]_0e^{-k_1t} + \epsilon_N[U]_0 \tag{A-8}$$

Only the reaction from U to I, which is a single-exponential process, can be observed experimentally. The refolding amplitude is $(\epsilon_U - \epsilon_N)[U]_0$, similar to that of U_{vf}. On the other hand, for U_s^I, $\epsilon_I = \epsilon_U$ as shown in eq 12. Hence, we can write for U_s^I

$$\text{total absorbance}(t) = (\epsilon_U - \epsilon_N)[U]_0\{[k_2/(k_2 - k_1)]e^{-k_1t} - [k_1/(k_2 - k_1)]e^{-k_2t}\} + \epsilon_N[U]_0 \tag{A-9}$$

If the proline isomerization step from U_s^I to I′ is much slower than the conformational folding process from I′ to N (i.e. if $k_1 \ll k_2$) under favorable folding conditions, then eq A-9 reduces to

$$\text{total absorbance}(t) = (\epsilon_U - \epsilon_N)[U]_0e^{-k_1t} + \epsilon_N[U]_0 \tag{A-10}$$

Only the slow isomerization process is observed experimentally, and the refolding amplitude is $(\epsilon_U - \epsilon_N)[U]_0$.

U_s^{II}. The refolding pathway of U_s^{II} is given by a four-state sequential folding model (eq 10d). The refolding reaction is

$$U_s^{II} \xrightarrow{k_1} I_1 \xrightarrow{k_2} I_N(sII) \xrightarrow{k_3} N \tag{A-11}$$

The equation can be solved following the procedure of Benson (1960). We assume that, at zero refolding time, only U_s^{II} is present at a concentration of $[U_s^{II}]_0$. The following expressions can be derived:

$$[U_s^{II}] = [U_s^{II}]_0e^{-k_1t} \tag{A-12a}$$

$$[I_1] = [U_s^{II}]_0[k_1/(k_2 - k_1)](e^{-k_1t} - e^{-k_2t}) \tag{A-12b}$$

$$[I_N(sII)] = [U_s^{II}]_0\{k_1k_2/[(k_2 - k_1)(k_3 - k_1)]\}e^{-k_1t} - [U_s^{II}]_0\{k_1k_2/[(k_2 - k_1)(k_3 - k_2)]\}e^{-k_2t} + [U_s^{II}]_0\{k_1k_2/[(k_3 - k_1)(k_3 - k_2)]\}e^{-k_3t} \tag{A-12c}$$

$$[N] = [U_s^{II}]_0 - [U_s^{II}]_0\{k_2k_3/[(k_2 - k_1)(k_3 - k_1)]\}e^{-k_1t} + [U_s^{II}]_0\{k_1k_3/[(k_2 - k_1)(k_3 - k_2)]\}e^{-k_2t} - [U_s^{II}]_0\{k_1k_2/[(k_3 - k_1)(k_3 - k_2)]\}e^{-k_3t} \tag{A-12d}$$

From eqs 11 and 12, $\epsilon_U = \epsilon(U_s^{II}) = \epsilon(I_1)$ and $\epsilon_N = \epsilon[I_N(sII)]$. Therefore, the total absorbance change as a function of the refolding time is given by

$$\text{total absorbance}(t) = \epsilon_U[U_s^{II}] + \epsilon_U[I_1] + \epsilon_N[I_N(sII)] + \epsilon_N[N] \tag{A-13}$$

Substitution of eqs A-12a−d into eq A-13 results in

$$\text{total absorbance}(t) = (\epsilon_U - \epsilon_N)[U_s^{II}]_0\{[k_2/(k_2 - k_1)]e^{-k_1t} - [k_1/(k_2 - k_1)]e^{-k_2t}\} + \epsilon_N[U_s^{II}]_0 \tag{A-14}$$

which is the same as eq A-9. This is expected because we cannot distinguish between N and $I_N(sII)$ by absorbance. Although the relaxation rate from U_s^{II} to I_1 is not well-

determined, nevertheless, it is expected to be much faster than the rate of formation of I_N(sII) (Schmid & Baldwin, 1979b; Udgaonkar & Baldwin, 1990). Hence, to a first approximation, $k_1 \gg k_2$. Consequently, eq A-14 reduces to

$$\text{total absorbance(t)} = (\epsilon_U - \epsilon_N)[U_s^{II}]_0 e^{-k_2 t} + \epsilon_N[U_s^{II}]_0 \quad \text{(A-15)}$$

The refolding process observed experimentally corresponds to the formation of I_N(sII) from I_1. The amplitude obtained is $(\epsilon_U - \epsilon_N)[U_s^{II}]_0$.

Therefore, upon the refolding of the different unfolded species, the refolding process for each species would appear to be a single-exponential process, and the amplitude obtained for the refolding decay curve is equal to the concentration of that species at zero refolding time multiplied by $(\epsilon_U - \epsilon_N)$. In other words, $\alpha_{vf}:\alpha_f:\alpha_m:\alpha_{sII}:\alpha_{sI} = [U_{vf}]_0:[U_f]_0:[U_m]_0:[U_s^{II}]_0:[U_s^{I}]_0$, where α_i is the absorbance-monitored refolding amplitude of species i.

REFERENCES

Adler, M., & Scheraga, H. A. (1990) *Biochemistry 29*, 8211−8216.
Anderson, D. G., Hammes, G. G., & Walz, F. G., Jr. (1968) *Biochemistry 7*, 1637−1645.
Antosiewicz, J., McCammon, J. A., & Gilson, M. K. (1994) *J. Mol. Biol. 238*, 415−436.
Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M., & Fersht, A. R. (1995) *J. Mol. Biol. 254*, 305−321.
Baker, W. R., & Kintanar, A. (1996) *Arch. Biochem. Biophys. 327*, 189−199.
Beaven, G. H., Holiday, E. R., & Johnson, E. A. (1955) in *The Nucleic Acids: Chemistry and Biology* (Chargaff, E., & Davidson, J. N., Eds.) Vol. 1, pp 493−553, Academic Press, New York.
Benson, S. W. (1960) *The Foundations of Chemical Kinetics*, Chapter 3, pp 39−42, McGraw-Hill Book Co., New York.
Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
Brems, D. N., & Baldwin, R. L. (1985) *Biochemistry 24*, 1689−1693.
Buckler, D. R., Haas, E., & Scheraga, H. A. (1995) *Biochemistry 34*, 15965−15978.
Chavez, L. G., Jr., & Scheraga, H. A. (1980) *Biochemistry 19*, 1005−1012.
Cook, K. H., Schmid, F. X., & Baldwin, R. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6157−6161.
Denton, J. B., Konishi, Y., & Scheraga, H. A. (1982) *Biochemistry 21*, 5155−5163.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
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.
Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., & Wright, P. E. (1988) *J. Mol. Biol. 201*, 161−200.
Garel, J.-R., & Baldwin, R. L. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3347−3351.
Garel, J.-R., Nall, B. T., & Baldwin, R. L. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 1853−1857.
Grathwohl, C., & Wuthrich, K. (1976) *Biopolymers 15*, 2025−2041.
Henkens, R. W., Gerber, A. D., Cooper, M. R., & Herzog, W. R., Jr. (1980) *J. Biol. Chem. 255*, 7075−7078.
Hermans, J., Jr., & Scheraga, H. A. (1961) *J. Am. Chem. Soc. 83*, 3293−3300.
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.
Hummel, J. P., Ver Ploeg, D. A., & Nelson, C. A. (1961) *J. Biol. Chem. 236*, 3168−3172.
Kiefhaber, T., & Schmid, F. X. (1992) *J. Mol. Biol. 224*, 231−240.
Kim, P. S., & Baldwin, R. L. (1980) *Biochemistry 19*, 6124−6129.
Krebs, H., Schmid, F. X., & Jaenicke, R. (1983) *J. Mol. Biol. 169*, 619−635.
Laskowski, M., Jr., & Scheraga, H. A. (1954) *J. Am. Chem. Soc. 76*, 6305−6319.
Lin, L.-N., & Brandts, J. F. (1983) *Biochemistry 22*, 564−573.
Lin, L.-N., & Brandts, J. F. (1987) *Biochemistry 26*, 3537−3543.
Liu, W., & Tsou, C.-L. (1987) *Biochim. Biophys. Acta 916*, 465−473.
Logan, T. M., Thériault, Y., & Fesik, S. W. (1994) *J. Mol. Biol. 236*, 637−648.
Mui, P. W., Konishi, Y., & Scheraga, H. A. (1985) *Biochemistry 24*, 4481−4489.
Nall, B. T., Garel, J.-R., & Baldwin, R. L. (1978) *J. Mol. Biol. 118*, 317−330.
Neri, D., Billeter, M., Wider, G., & Wüthrich, K. (1992) *Science 257*, 1559−1563.
Nozaki, Y. (1972) *Methods Enzymol. 26*, 43−50.
Rothwarf, D. M., & Scheraga, H. A. (1993) *Biochemistry 32*, 2671−2679.
Schmid, F. X. (1981) *Eur. J. Biochem. 114*, 105−109.
Schmid, F. X. (1983) *Biochemistry 22*, 4690−4696.
Schmid, F. X., & Baldwin, R. L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4764−4768.
Schmid, F. X., & Baldwin, R. L. (1979a) *J. Mol. Biol. 133*, 285−287.
Schmid, F. X., & Baldwin, R. L. (1979b) *J. Mol. Biol. 135*, 199−215.
Schmid, F. X., & Blaschek, H. (1981) *Eur. J. Biochem. 114*, 111−117.
Schmid, F. X., Grafl, R., Wrba, A., & Beintema, J. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 872−876.
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.
Shortle, D. (1993) *Curr. Opin. Struct. Biol. 3*, 66−74.
Shortle, D. R. (1996) *Curr. Opin. Struct. Biol. 6*, 24−30.
Stimson, E. R., Montelione, G. T., Meinwald, Y. C., Rudolph, R. K. E., & Scheraga, H. A. (1982) *Biochemistry 21*, 5252−5262.
Swadesh, J. K., Montelione, G. T., Thannhauser, T. W., & Scheraga, H. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4606−4610.
Szabo, Z. G. (1969) in *Comprehensive Chemical Kinetics* (Bamford, C. H., & Tipper, C. F. H., Eds.) Vol. 2, pp 1−80, Elsevier, Amsterdam.
Tanford, C., & Hauenstein, J. D. (1956) *J. Am. Chem. Soc. 78*, 5287−5291.
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.
Voet, D., Gratzer, W. B., Cox, R. A., & Doty, P. (1963) *Biopolymers 1*, 193−208.
Wlodawer, A., Svensson, L. A., Sjölin, L., & Gilliland, G. L. (1988) *Biochemistry 27*, 2705−2717.
Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E., & Dyson, H. J. (1994) *J. Mol. Biol. 243*, 736−753.
Ybe, J. A., & Kahn, P. C. (1994) *Protein Sci. 3*, 638−649.
Zhang, O., & Forman-Kay, J. D. (1995) *Biochemistry 34*, 6784−6794.

BI960745A