Folding Pathway of *Escherichia coli* Ribonuclease HI: A Circular Dichroism, Fluorescence, and NMR Study

Kazuhiko Yamasaki,*,‡ Kyoko Ogasahara,§ Katsuhide Yutani,§ Motohisa Oobatake,‡ and Shigenori Kanaya‡,||

Protein Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565, Japan, and Institute for Protein Research, Osaka University, 3-2, Yamadaoka, Suita, Osaka 565, Japan

Received June 16, 1995; Revised Manuscript Received September 15, 1995[⊗]

ABSTRACT: The unfolding and refolding processes of *Escherichia coli* ribonuclease HI at 25 °C, induced by concentration jumps of either guanidine hydrochloride (GuHCl) or urea, were investigated using stopped-flow circular dichroism (CD), stopped-flow fluorescence, and NMR spectroscopies. Only a single exponential process was detected for the fast time scale unfolding (rate constants from 0.014 to 0.54 s^{-1}, depending on the final denaturant concentration). For refolding, the far-UV CD value largely recovered within 50 ms of the stopped-flow mixing dead time (burst phase). This phase was followed by either one or two phases, with rate constants from 0.035 to 2.45 s^{-1} as detected by CD and fluorescence, respectively. Although this protein has a single *cis*-Pro residue, a very slow phase due to proline isomerization was not observed, for either unfolding or refolding. The difference in the amplitudes of the burst phases for refolding in the far- and near-UV CD spectra revealed that an intermediate state exists, with the characteristics of a molten globule. Because the one-phased fast exponential process detected by CD corresponds to the slower of the two phases detected by fluorescence, the intermediate detected by CD might be the most stable. GuHCl denaturation experiments revealed that this intermediate cooperatively unfolds, with a transition midpoint of 1.33 ± 0.03 M. The Gibbs free energy difference (ΔG) between the intermediate and the unfolded states, under physiological conditions (25 °C, pH 5.5, and 0 M GuHCl), was estimated to be 20.0 ± 2.3 kJ mol^{-1}. Therefore, it is reasonable to assume that the refolding intermediate, rather than the unfolded state, is the latent denatured state under physiological conditions. Approximately linear relationships between the GuHCl concentration and the logarithm of the microscopic rate constants determined by CD and fluorescence were also observed. By extrapolation to a GuHCl concentration of 0 M, activation Gibbs free energies of 98.5 ± 1.1 kJ mol^{-1} for unfolding and 69.5 ± 0.2 kJ mol^{-1} for refolding under physiological conditions were obtained. The hydrogen-exchange−refolding competition combined with two-dimensional NMR revealed that the amide protons of α-helix I are the most highly protected, suggesting that α-helix I is the initial site of protein folding. The CD and NMR data showed that the intermediate state has a structure similar to that of the acid-denatured molten globule.

To better understand the molecular architecture of proteins, the folding pathway from the unfolded state to the native state should be well characterized. For several small globular proteins, transient intermediates have been found to accumulate during the folding process [reviewed in Kim and Baldwin (1990)]. Detection and characterization of the folding intermediates are the key steps in understanding the folding pathway. Stopped-flow circular dichroism (CD)[1] is the most popular technique to detect transient folding intermediates (Kuwajima et al., 1985). From the far- and near-ultraviolet (UV) regions of CD spectra, information about the secondary and tertiary structures, respectively, of the intermediates can be obtained. The pulsed hydrogen−deuterium exchange (H−D exchange) labeling method, combined with two-dimensional (2D) NMR (Udgaonkar & Baldwin, 1988; Roder et al., 1988), provides detailed information on the secondary structure formation. The hydrogen-exchange−refolding competition method (Schmid & Baldwin, 1979), combined with 2D NMR (Roder & Wüthrich, 1986; Miranker et al., 1991), is similar to and easier than the pulsed H−D exchange labeling. Using these methods, transient folding intermediates have been detected and characterized at an atomic resolution for various proteins, such as RNase A (Udgaonkar & Baldwin, 1988, 1990), cytochrome *c* (Roder et al., 1988), barnase (Bycroft et al., 1990), hen lysozyme (Miranker et al., 1991; Radford et al., 1992), ubiquitin (Briggs & Roder, 1992), T4 lysozyme (Lu & Dahlquist, 1992), interleukin-1β (Varley et al., 1993), RNase T1 (Mullins et al., 1993), apomyoglobin (Jennings & Wright, 1993), and apoplastocyanin (Koide et al., 1993).

Intermediates have also been found in the equilibrium unfolding process (Yutani et al., 1979; Kuwajima, 1989; Kim & Baldwin, 1990). They are predominantly observed under mild denaturation conditions, such as a low pH and a moderate concentration of denaturants. Most of these

* To whom correspondence should be addressed (telephone, +81-6-872-8204; Telefax, +81-6-872-8210; E-mail, yamasaki@pes4.peri.co.jp).

‡ Protein Engineering Research Institute.

§ Osaka University.

|| Present address: Department of Material and Life Science, Faculty of Engineering, Osaka University, 2-1, Yamadaoka, Suita, Osaka 565, Japan.

⊗ Abstract published in *Advance ACS Abstracts,* November 15, 1995.

[1] Abbreviations: ASA, accessible surface area; CD, circular dichroism; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; ΔG, Gibbs free energy difference; GuHCl, guanidine hydrochloride; H−D exchange, hydrogen−deuterium exchange; HSQC, heteronuclear single-quantum coherence spectroscopy; pD_{read}, the pH meter reading value of the D_2O buffer; RNase HI, ribonuclease HI; UV, ultraviolet.

0006-2960/95/0434-16552$09.00/0

intermediates have compact forms with a significant amount of secondary structure, but with less ordered tertiary structures. These features define the molten globule (Ohgushi & Wada, 1983). It has been suggested, from stopped-flow CD analyses, that the molten globules of α-lactalbumin (Ikeguchi et al., 1986) and tryptophan synthase α-subunit (Ogasahara & Yutani, 1994), induced by low concentrations of denaturants, are identical with the transient folding intermediates. Pulsed H−D exchange labeling with NMR has also revealed that the acid-induced molten globule of apomyoglobin is similar to the transient folding intermediate (Jennings & Wright, 1993).

Ribonuclease HI (RNase HI) from *Escherichia coli* enzymatically cleaves the RNA moiety of a DNA−RNA hybrid duplex (Crouch & Dirksen, 1982). This protein has a mass weight of 17.6 kDa (Kanaya & Crouch, 1983) and consists of five α-helices and five β-strands, as revealed by X-ray crystallography (Katayanagi et al., 1990, 1992; Yang et al., 1990). The assignments of the NMR signals, as well as the secondary structure determination in an aqueous solution, have been completed (Yamazaki et al., 1991, 1993). This protein is free of disulfide bonds (Kanaya et al., 1990; Katayanagi et al., 1990, 1992; Yang et al., 1990) and retains full enzymatic activity after the denaturation−renaturation process (Kanaya & Itaya, 1992; Kimura et al., 1992a). Therefore, *E. coli* RNase HI is suitable as a model protein for folding pathway analyses.

During the equilibrium unfolding by either urea or guanidine hydrochloride (GuHCl), no intermediate was found (Kanaya & Itaya, 1992; Kimura et al., 1992a,b). However, the acid-denatured states (A-states) of the wild type (Oobatake et al., 1993) and the cysteine-free variant (Dabora & Marqusee, 1994) of RNase HI have the characteristics of the molten globule, which could be a stable structure that is equivalent to the transient refolding intermediate.

In the present study, we analyzed the folding pathway of *E. coli* RNase HI using CD, aromatic fluorescence, and NMR. Far- and near-UV stopped-flow CD detected an early folding intermediate during the refolding process. The structure and stability of this intermediate were analyzed by far-UV CD and NMR.

MATERIALS AND METHODS

Preparation of RNase HI Sample. ^{15}N-Labeled and nonlabeled *E. coli* RNase HI were overproduced and purified as described previously (Yamasaki et al., 1995) by a slight modification of the method of Kanaya et al. (1993). The concentration of the purified protein was evaluated from the UV absorbance at 280 nm, based on the value of 2.02 for a 1 mg/mL solution (Kanaya et al., 1990). From a 1 L scale culture, 30−40 mg of nonlabeled RNase HI or 15−25 mg of ^{15}N-labeled RNase HI was obtained. Nonlabeled proteins were used for CD and fluorescence, and the ^{15}N-labeled proteins were used for NMR. The proteins were dissolved in the buffers described below.

CD and Fluorescence Measurements. Stopped-flow and manual-mixing CD were carried out on a Jasco J-720 spectropolarimeter, as described by Ogasahara and Yutani (1994), at 25 °C, pH 5.5. Stopped-flow aromatic fluorescence was measured on a Photal RA-401 (Ohtsuka Electronics), also as described by Ogasahara and Yutani (1994), at 25 °C, pH 5.5. For far-UV CD and aromatic fluorescence, the proteins (1.0 mg/mL) were equilibrated with acetate buffer [10 mM CH_3COONa (pH 5.5)] in the presence (for the refolding analysis) or absence (for the unfolding analysis) of 3.3 M GuHCl (Nacalai Pure Chemicals Co. Ltd.) using a Centricon 10 filtration unit (Amicon). The reaction was triggered by concentration jumps of GuHCl, which were carried out by mixing the protein solution with a 10-fold greater volume of acetate buffer containing various concentrations of GuHCl (0−3.3 M). For near-UV CD, a 10-fold higher concentration of the protein was used. Far- and near-UV CD were measured at 220 and 292 nm, respectively, as functions of time. Aromatic fluorescence was collected in the range over 300 nm using a UV30 cut filter, with excitation at 280 nm. The lengths of the cells for stopped-flow CD, manual-mixing CD, and stopped-flow fluorescence were 10, 10, and 2 mm, respectively. The far-UV CD spectra of the acid-denatured states (10 °C) were measured at a protein concentration of 0.125 mg/mL in 10 mM glycine hydrochloride buffer (pH 1.0) using a 2 mm cell.

Analyses of CD and Fluorescence Data. The mean residue ellipticity, θ, which has the units of deg cm^2 dmol^{-1}, was calculated using an average amino acid molecular weight of 110. The CD and fluorescence time courses were fitted to exponential curves using the equation

$$A(t) - A(\infty) = \sum_{i=1}^{N} A_i \exp(-k_i t) \quad (1)$$

where $A(t)$ is the value of the CD or fluorescence intensity, $A(\infty)$ represents those values at the steady-state level, and A_i and k_i are the amplitude and the apparent rate constant, respectively, of the ith kinetic phase. The data analyses were performed on an NEC personal computer, using software provided by Dr. Kuwajima (Sugawara et al., 1991).

The dependence of the CD values of the intermediate on the GuHCl concentration was fitted to the equation for the two-state mechanism

$$\theta = \theta_i f_i + \theta_u f_u \quad (2)$$

where θ is the ellipticity value depending on the GuHCl concentration, θ_i and θ_u are those of the intermediate and unfolded states, respectively, and f_i and f_u are their fractions at equilibrium ($f_i + f_u = 1$). The equilibrium constant $K_{IU} = f_u/f_i$ can be expressed as

$$K_{IU} = \exp(-\Delta G_{IU}/RT) \quad (3)$$

where ΔG_{IU} is the Gibbs free energy difference between the unfolded and intermediate states, R is the gas constant, and T is the temperature (298 K). The dependence of θ_u, θ_i, and ΔG_{IU} on the GuHCl concentration [D] was assumed to be linear (Pace, 1990): for ΔG

$$\Delta G_{IU} = \Delta G_{IU}^{H_2O} + m_{IU}[D] \quad (4)$$

where $\Delta G_{IU}^{H_2O}$ is the ΔG_{IU} value in the absence of GuHCl (physiological condition) and m_{IU} is the coefficient for the dependence of ΔG_{IU} on the GuHCl concentration. The value θ_u was determined in advance by linear fitting of the unfolded region (more than 2.5 M GuHCl) of the equilibrium unfolding data by Kimura et al. (1992b). Nonlinear least-squares minimization of the difference between the observed and calculated ellipticity values ($\theta_{obs} - \theta_{calc}$) was done with

an in-house FORTRAN77 program employing the algorithm of Levenberg−Marquardt's method (Levenberg, 1944; Marquardt, 1963) in an IMSL library (IMSL Inc.) on a VAX computer (DEC). From this procedure, four parameters, $\Delta G_{IU}^{H_2O}$, m_{IU}, the θ_i value at 0 M GuHCl, and its dependence on the GuHCl concentration, were obtained. To estimate the uncertainty of these parameters, a Monte Carlo simulation (Kamath & Shriver, 1989; Palmer et al., 1991) was done using an algorithm in the IMSL library, as follows. With Gaussian randomization of the ellipticity values at each time point, the minimization described above was repeated 2000 times. For the standard deviation in the Gaussian distribution, the root mean square of the noise in the CD measurement was employed. The resultant average and the standard deviation were adopted as the final value and the uncertainty, respectively.

The equilibrium unfolding data by Kimura et al. (1992b) were fitted to the equation for a three-state equilibrium

$$\theta = \theta_n f_n + \theta_i f_i + \theta_u f_u \quad (5)$$

where θ_n, θ_i, and θ_u are the ellipticity of the native, intermediate, and unfolded states, respectively, and f_n, f_i, and f_u are their fractions at equilibrium ($f_n + f_i + f_u = 1$). The value θ_i and equilibrium constant $K_{IU} = f_u/f_i$ determined above were employed. θ_n, θ_u, and $\Delta G_{NU} = -RT \ln(f_u/f_n)$ were determined as linear functions of GuHCl concentration by minimization of $[\theta_{obs} - \theta_{calc}]$. To estimate the uncertainty, the root mean square of the fitting residuals of $[\theta_{obs} - \theta_{calc}]$ was employed for the standard deviation in the Monte Carlo simulation.

By assuming a three-state mechanism with a rapid equilibrium between the intermediate and unfolded states, the apparent rate constants were converted to the microscopic rate constants using the equations by Kuwajima et al. (1989):

$$k_f = \frac{1 + K_{IU}}{1 + K_{NI} + K_{NU}} k_{app} \quad (6)$$

$$k_u = \frac{K_{NI} + K_{NU}}{1 + K_{NI} + K_{NU}} k_{app} \quad (7)$$

where k_f and k_u are the microscopic rate constants for refolding and unfolding, respectively, k_{app} is the apparent rate constant, and $K_{IU} = f_u/f_i$, $K_{NI} = f_i/f_n$, and $K_{NU} = f_u/f_n$ are the equilibrium constants. The dependence of the logarithm of the microscopic rate constants on the GuHCl concentration [D] was fitted to a linear line:

$$\ln k = \ln k^{H_2O} - m^{\ddagger}[D]/RT \quad (8)$$

where k is the microscopic rate constant (k_f or k_u), k^{H_2O} is that under physiological conditions ($k_f^{H_2O}$ or $k_u^{H_2O}$), and $m^{\ddagger}$ is the coefficient for the dependence ($m_f^{\ddagger}$ or $m_u^{\ddagger}$). For estimation of the uncertainty, the Monte Carlo simulation was employed in the same way as described above, using the root mean square of the fitting residuals as the standard deviation.

Hydrogen-Exchange Competition. The H−D exchange−refolding competition method (Schmid & Baldwin, 1979; Roder & Wüthrich, 1986; Miranker et al., 1991) was employed to determine the fast folding sites in comparison with the hydrogen-exchange rates. This method was applied at three different pH values (5.5, 6.5, and 7.5) to adjust the hydrogen-exchange rate. For pH 5.5, 10 mM acetate buffer made from CD_3COOD (CIL) and NaOH was used. For pH 7.5 and 6.5, 10 mM phosphate buffer was used. The buffers made with D_2O were designated as D_2O buffers to distinguish them from the buffers made with H_2O (H_2O buffers). For the D_2O buffers, the pH values were not corrected for isotope effects: pD_{read}, representing the pH meter reading value of the D_2O buffer, was employed. Four milligrams of RNase HI was dissolved in 200 μL of the H_2O buffers containing 7.0 M urea (Wako Pure Chemicals Co. Ltd.). This protein solution (200 μL) and the D_2O buffer (5 mL), at the corresponding pD_{read} without urea, were preincubated at 25 °C for more than 5 min. The competition between the refolding and the H−D exchange reactions was initiated by mixing these two solutions in a very short time (within 500 ms). After 1 min, the H−D exchange reaction was quenched by ice cooling and pH adjustment by dilution with 10 mL of a D_2O buffer [250 mM CD_3COONa (pD_{read} 5.5)]. Within 5 h, the volumes of the refolded protein solutions were reduced to 200 μL and were equilibrated with the D_2O buffer [100 mM CD_3COONa (pD_{read} 5.5)] using a Centricon 10 unit. This procedure was done at 4 °C to minimize the amide H−D exchange during the process. The samples were maintained in the NMR probe at 27 °C for 30 min before the initiation of the NMR measurements.

For a control sample without the competition process, 4 mg of RNase HI was equilibrated with H_2O and lyophilized. The lyophilysate was dissolved in the D_2O buffer [100 mM CD_3COONa (pD_{read} 5.5)] and incubated at 27 °C for 30 min before the NMR measurement.

NMR Measurements and Data Processing. All NMR spectra were measured on a Bruker AM-500 spectrometer (500.13 MHz for ^1H and 50.7 M for ^{15}N) at a probe temperature of 27 °C. Proton and nitrogen chemical shifts were determined relative to sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) and liquid NH_3, respectively. Two-dimensional heteronuclear single-quantum coherence spectroscopy (HSQC) (Bodenhausen & Ruben, 1980) was carried out with a delay of 2.25 ms. The sweep width for ^1H was 7042 Hz, and that for ^{15}N was 1830 Hz. GARP1 ^{15}N decoupling (Shaka et al., 1985) was employed during the detection period. Free induction decays (48 scans) of 2K data points in the t_2 domain were collected for 200 data points in the t_1 domain using time-proportional phase incrementation (Marion & Wüthrich, 1983). Data processing was done with the FELIX 2.30 software (Biosym) on an IRIS indigo II computer (Silicon Graphics). With squared sine-bell windows of 60° for the t_2 and 90° for the t_1 dimensions, followed by zero filling and Fourier transformation, spectra of 2048 × 512 data points were obtained. The cross-peak intensities of the refolded proteins after the competition processes were measured and compared with those of the control sample. To normalize the intensities, those of the δ-methyl proton of Ile116 in one-dimensional NMR, measured just before each 2D HSQC, were used as the standard.

Estimation of Intrinsic H−D Exchange Rates of RNase HI. The intrinsic H−D exchange rates of individual amino acid residues were obtained by the method of Bai et al. (1993) at pD_{read} 5.5, 6.5, and 7.5 (to be corrected by $pD = pD_{read} + 0.4$; Glasoe & Long, 1960). This method distinguishes the charges of the Asp, Glu, and His residues, the disulfide bonds of the Cys residues, and the *cis−trans*

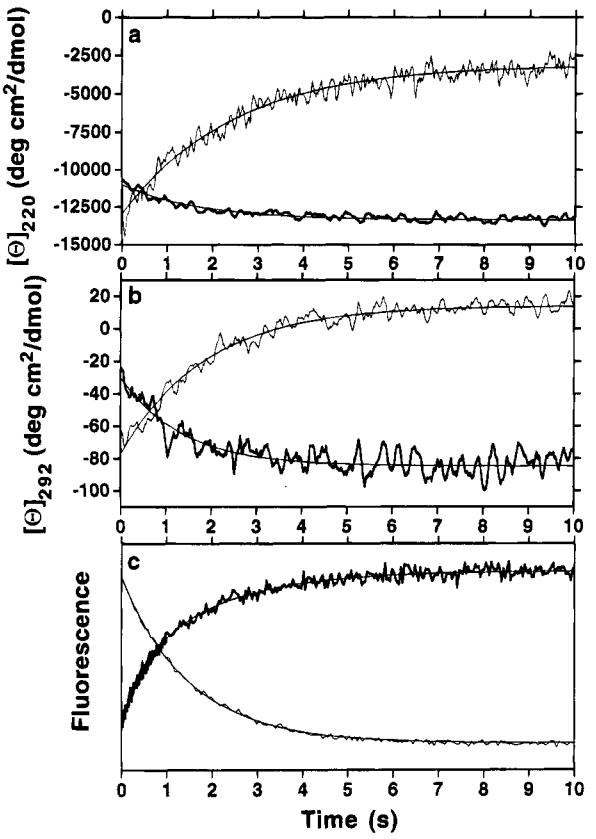

FIGURE 1: Kinetic curves of unfolding (thin line) and refolding (thick line) of *E. coli* RNase HI as monitored by far-UV CD at 220 nm (a), near-UV CD at 292 nm (b), and aromatic fluorescence over 300 nm with excitation at 280 nm (c), using the stopped-flow method at pH 5.5 and 25 °C. The final GuHCl concentrations were 3.0 M for unfolding and 0.9 M for refolding. Theoretical best-fit exponential curves are also indicated. Those with single phases were employed, except for the refolding monitored by fluorescence (two phases). For the refolding monitored by fluorescence, a curve with a shorter scale (1.0 s) was superimposed to improve the resolution. In this figure, the apparent kinetic rate constants for the unfolding are 0.42 s^{-1} by far-UV CD (a), 0.54 s^{-1} by near-UV CD (b), and 0.67 s^{-1} by fluorescence (c). Those for refolding are 0.59 s^{-1} by far-UV CD (a), 0.74 s^{-1} by near-UV CD (b), and two phases of 0.51 and 1.95 s^{-1} by fluorescence (c). The amplitude ratio of the slower to faster phases in the refolding monitored by the fluorescence (c) is about 13:7. It should be noted that the scales for the two curves in the fluorescence intensity are arbitrary.

isomerization of the Pro residues. We assumed a random-coiled chain for these residues as follows. According to the pK_a values for model compounds [see Creighton (1993), for example], Asp and Glu residues are charged at the three pDs, and His residues are charged at pD 5.9 but not at pD 6.9 and 7.9. The Cys residues do not form the disulfide bonds and the Pro residues have the *trans* conformation.

RESULTS

Unfolding and Refolding Kinetics Monitored by CD and Fluorescence. Unfolding and refolding reactions were initiated by GuHCl concentration jumps from 0 and 3.3 M, respectively, to various concentrations above 0.3 M and below 3.0 M, respectively, at pH 5.5 and 25 °C. The concentration jumps were triggered by stopped-flow mixing. For the relatively slow reaction, manual mixing, which is more sensitive, was also employed. Figure 1 shows typical kinetic progress curves. The CD at 220 nm (far-UV CD) reflects mostly changes in the secondary structure, as shown in Figure 1a. For the unfolding kinetics, after the reaction was triggered, the negative far-UV CD decreased from that of the native state to that of the unfolded state. This progression was fitted to a single exponential curve in eq 1. The observed amplitude approached 94% of the difference between the CD values of the native and unfolded proteins. Therefore, the two-state mechanism, without an intermediate, seems to be adopted for the unfolding process of RNase HI. In contrast, for the refolding, the CD value at 220 nm sharply recovered to 83% within the dead time (50 ms) of the stopped-flow mixing (Figure 1a). After this burst phase, the folding kinetics followed in a single exponential manner toward the value of the native state. This suggests that a substantial amount of secondary structure is formed at a very early stage of the refolding reaction.

The disruption and formation of the tertiary structure were monitored by the CD at 292 nm (near-UV CD) and the aromatic fluorescence over 300 nm with excitation at 280 nm (Figure 1b,c). The kinetic progress for the unfolding and the refolding monitored by the near-UV CD could also be fitted to single exponential curves (Figure 1b). The amplitudes of the exponential phases were 92% for the unfolding and 56% for the refolding. The fluorescence also showed a single exponential curve for the unfolding (Figure 1c). For the refolding monitored by the fluorescence, however, the fitting was slightly better with double exponential curves (Figure 1c).

These refolding and unfolding kinetics data by far-UV CD and fluorescence were collected at various final GuHCl concentrations, as shown in Figure 2. Although the GuHCl concentration strongly affected the apparent rate constants for unfolding and refolding, the trends stated above were valid at all concentrations. In the far-UV CD, the kinetic amplitude was close to 100% for the unfolding, while a much smaller amplitude was observed after the burst recovery of the ellipticity for the refolding. The kinetic curves monitored by fluorescence could be fitted to a single exponential curve for the unfolding but fit slightly better with a double exponential curve for the refolding.

It should be noted that, in the refolding kinetics, a much smaller burst phase was observed by the near-UV than by the far-UV CD (Figure 1a,b). Therefore, this proves that this burst phase corresponds to the formation of an intermediate with a considerable amount of secondary structure (about 80% of the native state) but with a less-ordered tertiary structure. This is characteristic of a molten globule (Ohgushi & Wada, 1983; Kuwajima, 1989; Ptitsyn et al., 1990). The apparent kinetic rates for the unfolding and refolding obtained by the near-UV CD and the fluorescence (the slower phase for the refolding) were similar to those obtained by the far-UV CD at any GuHCl concentration (Figures 1 and 2), which suggests the simultaneous formation of the secondary and tertiary structures after the burst phase.

Although the single exponential curves in both the far- and near-UV CD suggested that only one intermediate can exist, the fluorescence data indicate that at least two intermediates exist, due to the double exponential curves. However, the major and slower phase of the transition by fluorescence corresponds to that by the far- and near-UV CD. The minor phase could not be detected by CD, probably because of poor sensitivity or because the CD values, but not the fluorescence intensities, of the intermediates are similar to each other. The kinetic phase detected by CD

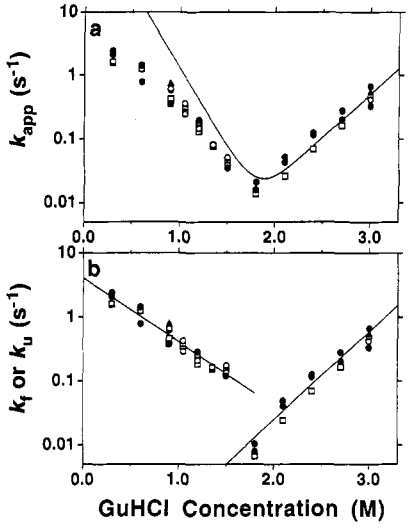

FIGURE 2: Dependence of the apparent (a, k_{app}) and microscopic (b, k_f or k_u) rate constants on the GuHCl concentration at pH 5.5 and 25 °C. Data were collected by far-UV CD (220 nm) with stopped-flow mixing (○) and with manual mixing (□), by near-UV CD (292 nm) with stopped-flow mixing (▲), and by aromatic fluorescence with stopped-flow mixing (●). Data in the region with concentrations of GuHCl lower than 1.6 M are the refolding rate constants, and those in the region higher than 1.6 M are the unfolding ones. The microscopic rate constants for the refolding and unfolding are obtained with eqs 6 and 7, respectively, assuming a three-state mechanism with rapid exchange between the intermediate and unfolded states. For the refolding monitored by fluorescence, the rates of the major slower phase are plotted. The curve in (a) shows the fitting of the unfolding rate constants assuming a two-state mechanism ($k_{app} = k_f + k_u$) with linear dependence of the logarithm of k_u on the GuHCl concentration. The lines in (b) represent the best-fit linear lines for the logarithms of k_f and k_u.

might represent the most stable intermediate species. Therefore, in this paper, we tentatively assume a single intermediate, for simplicity. Thus, the early refolding intermediate was detected, while no intermediate was detected in the unfolding process.

Dependence of CD Values Immediately after the Initial Burst Phase in the Refolding Kinetics on the GuHCl Concentration. Figure 3 shows the CD values at 220 nm immediately after the burst recovery as a function of the GuHCl concentration, together with the equilibrium unfolding transition curve previously reported (Kimura et al., 1992b). When the GuHCl concentration is lower than 1.0 M, the CD values showed little dependence on the concentration (Figure 3). With concentrations higher than 1.0 M, however, the CD values of the initial points increased with a sigmoidal curve, like the equilibrium unfolding of the native state. The sigmoidal curve could be fitted well to eqs 2−4, which shows the cooperative two-state transition (Pace et al., 1990). This suggests that there is a quick equilibrium between the unfolded and the intermediate states within the stopped-flow dead time (50 ms) and that the balance between them shifts to the unfolded state when the concentration of GuHCl is increased. The fitting indicated that the Gibbs free energy difference between the intermediate and the unfolded states in the absence of GuHCl (under "physiological" conditions) ($\Delta G_{IU}^{H_2O}$) is 20.0 ± 2.3 kJ mol^{-1} and that the dependence of the ΔG value on the GuHCl concentration (m_{IU}) is −15.0 ± 1.5 kJ mol^{-1} M^{-1}. The GuHCl concentration at the transition midpoint ($[D]_{1/2}$) was 1.33 ± 0.03 M.

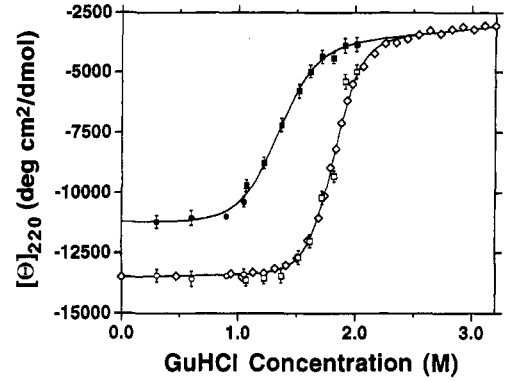

FIGURE 3: Dependence of far-UV CD values at 220 nm on GuHCl concentration at pH 5.5 and 25 °C. The CD values immediately after the burst phase (● by stopped-flow mixing; ■ by manual mixing) and those for the plateau level (○ by stopped-flow mixing; □ by manual mixing) of the kinetic CD progress are indicated, together with the noise levels, shown by error bars. The CD values obtained by the equilibrium unfolding experiment (Kimura et al., 1992b) are shown with (◇). The best-fit curve of the two-state analysis (eq 2) for (●) and (□) and that of the three-state analysis (eq 5) for (◇) are indicated as lines.

Employing these values, the previous data of the GuHCl-induced equilibrium unfolding, obtained by Kimura et al. (1992b), were reexamined assuming the three-state mechanism (eq 5). The determined ΔG^{H_2O}, m, and $[D]_{1/2}$ values for the native unfolded transition are 39.6 ± 1.5 kJ mol^{-1}, −21.9 ± 0.8 kJ mol^{-1} M^{-1}, and 1.81 ± 0.004 M, respectively, which are similar to the previously reported values (38.3 kJ mol^{-1}, −21.0 kJ mol^{-1} M^{-1}, and 1.83 M, respectively) obtained by assuming the two-state mechanism (Kimura et al., 1992b). This is because the intermediate has only a small contribution in the equilibrium unfolding transition. In fact, using the two equilibrium constants, the maximum fraction of the intermediate state in the equilibrium is calculated to be only 2.9%, at the GuHCl concentration of 1.72 M. Thus, the protein in the intermediate state is less stable than that in the native state, by 19.6 kJ mol^{-1} in ΔG^{H_2O} and by 0.48 M in $[D]_{1/2}$.

Activation Free Energies of Unfolding and Refolding. The final GuHCl concentration strongly affects the apparent rate constants for the unfolding and refolding (Figure 2a). In parallel with the increase in the GuHCl concentration, the unfolding rates increased and the refolding rates decreased. Apparent rate constants for the unfolding ranged from 0.014 to 0.54 s^{-1} in the GuHCl concentration range of 1.8−3.0 M, and those for the refolding ranged from 0.035 to 2.45 s^{-1} (slow phase for the aromatic fluorescence) in the GuHCl concentration range of 0.3−1.5 M. The relationship is characterized by a V-shape, with a minimum of about 1.8 M, which is equivalent to the midpoint of the transition curve from the native to the unfolded states (Figure 3). However, the fitting of the rate constants, assuming the two-state mechanism, was not successful, as shown in Figure 2a. This is similar to other proteins, such as barnase (Matouschek et al., 1990). Therefore, we assumed the three-state mechanism with rapid exchange between the intermediate and unfolded states, where the microscopic rate for refolding (k_f) represents that from the intermediate to the native states, and that for unfolding (k_u) represents that from the native to the intermediate states. In this case, the apparent rate constants can be converted to microscopic rate constants for the refolding

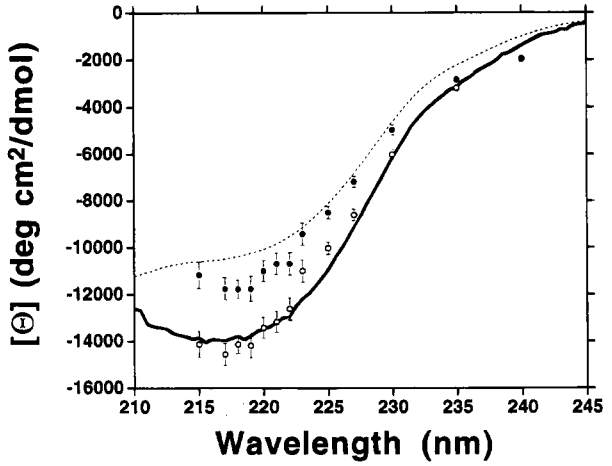

FIGURE 4: CD spectra of the early refolding intermediate state (●) together with the native state (pH 5.5, 25 °C, in 0.9 M GuHCl; thick solid line) and the acid-denatured molten globule (pH 1.0, 10 °C; thin dotted line). The CD spectrum for the intermediate was obtained from the CD values monitored at various wavelengths, immediately after the burst phase of the refolding triggered by a GuHCl concentration drop from 3 to 0.9 M. The plateau levels of the kinetic progressions are indicated by (○). The error bars show the noise levels.

and the unfolding with eqs 6 and 7 (Kuwajima et al., 1989). For both the unfolding and refolding, approximately linear relationships were observed between the denaturant concentration and the logarithm of the microscopic rate constants (Figure 2b). Especially for the refolding, the linearity is better for the microscopic rate constants than for the apparent ones. By extrapolation to 0 M GuHCl concentration, we estimated the rate constants under physiological conditions (k^{H_2O}) for the unfolding and refolding, using eq 8. The estimated k^{H_2O} were $3.7 \times 10^{-5} \pm 1.7 \times 10^{-5}$ s^{-1} for the unfolding and 4.1 ± 0.4 s^{-1} for the refolding.

We calculated the activation energy from these rate constants by Eyring's equation (Eyring, 1935)

$$\Delta G^{\ddagger H_2O} = RT \ln\left(\frac{k_B T}{h k^{H_2O}}\right) \quad (9)$$

where $\Delta G^{\ddagger H_2O}$ is the activation free energy under physiological conditions and k_B and h are the Boltzmann and Planck constants, respectively. The calculated $\Delta G^{\ddagger H_2O}$ values are 98.5 ± 1.1 kJ mol^{-1} for the unfolding from the native state to the unfolded state and 69.5 ± 0.2 kJ mol^{-1} for the refolding from the early folding intermediate state to the native state. The dependence of the activation energy on the GuHCl concentration ($m^{\ddagger}$) was -8.0 ± 0.4 kJ mol^{-1} M^{-1} for the unfolding ($m_u^{\ddagger}$) and 5.7 ± 0.2 kJ mol^{-1} M^{-1} for the refolding ($m_f^{\ddagger}$).

CD Spectra of the Early Refolding Intermediate. By collection of the CD values at various wavelengths immediately after the burst recovery in the kinetic refolding, the CD spectrum of the early folding intermediate state can be obtained. The refolding reactions were induced by a jump in the GuHCl concentration from 3.0 to 0.9 M at pH 5.5 and 25 °C. The wavelength was varied over the range of 215−240 nm. Figure 4 shows the far-UV CD spectrum of the refolding intermediate at the GuHCl concentration of 0.9 M, where the population of the unfolded state is negligible as compared with that of the intermediate. The spectra of the native and the acid-denatured molten globule states are also indicated. Although the negative CD values of the refolding intermediate are slightly larger than those of the acid-denatured molten globule, they are similar to each other. Both of the spectra showed ellipticity values that are 70−80% of that of the native state ($[\theta]_{220}$ is ca. −11 000 for the refolding intermediate and ca. −10000 for the acid-denatured state). Furthermore, the near-UV CD value at 292 nm immediately after the burst phase in the refolding kinetics ($[\theta]_{292}$ is ca. −30) and that of the acid-denatured molten globule ($[\theta]_{292}$ is ca. −24) are similar to each other. These results indicate that the structures of the two states have some similarity.

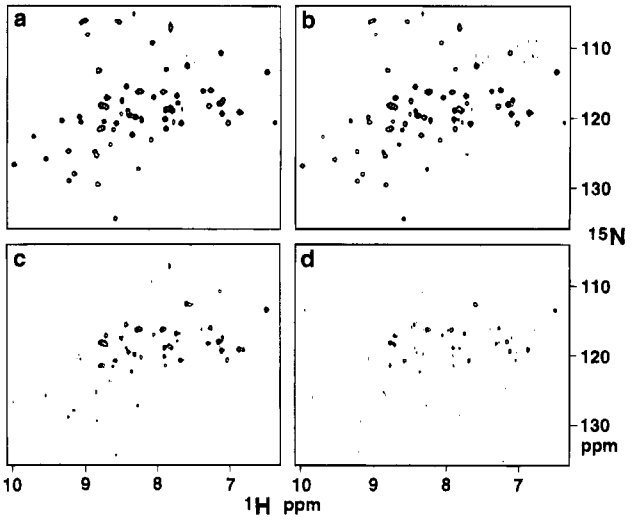

FIGURE 5: ^{1}H−^{15}N HSQC spectra of RNase HI at pH 5.5 and 27 °C: without competition (a) and after competition at pH 5.5 (b), pH 6.5 (c), and pH 7.5 (d).

Hydrogen Exchange−Refolding Competition. We found the existence of an early folding intermediate in the refolding of RNase HI. This intermediate has the features of a molten globule. To explore the secondary structure formed in the intermediate at an atomic resolution, we employed the hydrogen-exchange competition method (Schmid & Baldwin, 1979) combined with 2D NMR spectroscopy (Roder & Wüthrich, 1986; Miranker et al., 1991). In this method, the amide protons that form hydrogen bonds faster than the H−D exchange are protected from the exchange, and therefore, their NMR signals can be detected. Figure 5 shows the 2D ^{1}H−^{15}N HSQC spectra of the native proteins in a D_2O solution after the competition process (panels b−d) and without the process (control sample) (panel a). The assignments of the cross-peaks by Yamazaki et al. (1991) were easily applied to these spectra. Among the 149 backbone amide groups in RNase HI (the N-terminus and five Pro residues were excluded from the 155 residues), 59 groups were well protected from the hydrogen exchange in the native state (the protection factors are 10^3−10^7 at pD_{read} 5.5, 27 °C; K. Yamasaki, unpublished data) and were observed separately in D_2O solution (the cross-peaks of Val54 and Arg138 are partially overlapped). These protected residues are as follows: β-strand A, Val5, Glu6, Ile7, Phe8, Thr9, and Asp10; the turn between β-strands A and B, Gly18; β-strand B, Gly20, Gly21, Tyr22, Gly23, Ala24, Ile25, Leu26, and Arg27; β-strand C, Lys33, Phe35, Ala37, and Tyr39; the loop between β-strand C and α-helix I, Thr42 and Thr43; α-helix I, Arg46, Met47, Glu48, Leu49, Met50, Ala51, Ala52, Ile53, Val54, Ala55, Leu56, Ala58, and Leu59; β-strand D, Val65,

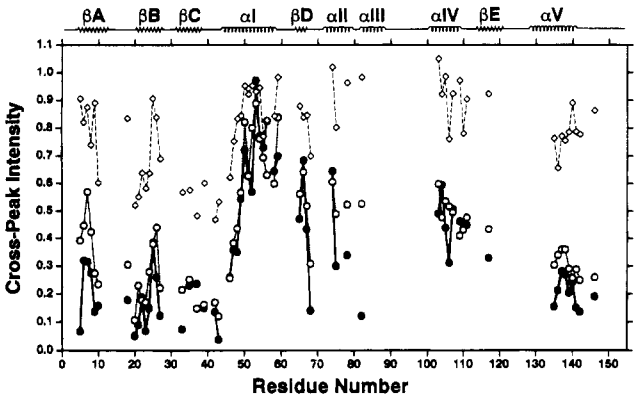

FIGURE 6: Ratios of cross-peak intensities after competition to those without competition as functions of residue number: pH 5.5 (◇ with broken lines), pH 6.5 (○ with dotted lines), and pH 7.5 (● with thick solid lines).

Ile66, Leu67, and Ser68; α-helix II, Val74, Arg75, and Ile78; α-helix IV, Leu103, Trp104, Gln105, Arg106, Leu107, Ala109, Ala110, and Leu111; β-strand E, Lys117; α-helix V, Glu135, Leu136, Ala137, Arg138, Ala139, Ala140, Ala141, and Met142; loop, Leu146. All of them serve as probes for the formation of hydrogen bonds, which mostly represent the formation of the secondary structure.

The H−D exchange during the competition process reduces the cross-peak intensities of the processed protein samples (Figures 5− d) relative to those of the intact (control) sample (Figure 5a). The intensity ratios of each residue in the samples after competition at pH 5.5, 6.5, and 7.5, relative to those of the control sample, are shown in Figure 6. The intrinsic H−D exchange rates of the unprotected peptides (intrinsic exchange rate) were calculated for RNase HI by the sequence-specific method of Bai et al. (1993). At pD_{read} 5.5 and 25 °C, the average of the intrinsic H−D exchange rates for the observed residues was 0.25 s^{-1}, and the time scale is 4 s. When the competition experiment was done under these conditions, the cross-peaks of most of the hydrogen-bonding amides were observed with the same intensities as those of the control. In contrast, for those in parts of the β-sheet, especially β-strand C (Lys33, Phe35, Ala37, and Tyr39), the cross-peak intensities decreased to about half of those of the control (Figures 5a,b and 6). This result indicates that the rate of hydrogen bond formation of most residues is much faster than the intrinsic H−D exchange rate (average 0.25 s^{-1}), but for the residues in β-strand C, the rates are comparable. It is likely that the formation of this part is the final step of the refolding process, which is possibly coupled with the formation of the tertiary structure.

At neutral pH, the intrinsic H−D exchange rate of the peptides increases by 10-fold as the pH increases by 1.0 unit (Bai et al., 1993). Therefore, it is expected that the cross-peak intensities would dramatically decrease when the pD_{read} is increased to 6.5 and further to 7.5 (Figures 5c,d and 6). At pD_{read} 6.5, where the time scale of the intrinsic H−D exchange is 400 ms (the average rate is 2.5 s^{-1}), the intensities of many cross-peaks decreased to less than half of that of the control sample (Figures 5a,c and 6). At pD_{read} 7.5, the time scale of the intrinsic H−D exchange is about 40 ms (the average rate is 25 s^{-1}), and the intensities of most of the cross-peaks further decreased (Figures 5d and 6). For some residues, however, the amide protons are still highly protected. Most of them are located in α-helix I in the native structure (Katayanagi et al., 1990, 1992; Yang et al., 1990), as shown in Figure 7. The intrinsic H−D exchange rates for the region are calculated to be only slightly lower (20 s^{-1}; average of residues from Leu49 to Leu59) than the average. Therefore, it is concluded that, in the refolding intermediate, some structure, consisting mainly of α-helix I, is already formed to protect the amide protons from the exchange. It should be noted that the residues surrounding α-helix I are also considerably protected (Figures 6 and 7).

Together with the CD results, the refolding of RNase HI can be summarized as follows. In a very short time after the initiation of the folding reaction, an intermediate is formed. This intermediate has the characteristics of a molten globule. In this molten globule, some structure, consisting mainly of α-helix I, is formed. After the formation of this intermediate structure, the remaining secondary structure and the tertiary structure are formed simultaneously.

DISCUSSION

Simple Kinetic Processes. It should be emphasized that all of the kinetic processes for protein unfolding and refolding can be fitted to simple (nearly single) exponential curves. For some proteins, such as RNase A and RNase T1, additional very slow phases (kinetic rates less than 10^{-3} s^{-1}) have been observed, which are thought to be due to proline *cis*−*trans* isomerization (Garel & Baldwin, 1973; Brandts et al., 1975; Kiefhaber et al., 1990a). For several proteins, replacement of the *cis*-proline(s) has eliminated this very slow phase and has simplified the kinetic progress (Kelley & Richards, 1987; Kiefhaber et al., 1990b; Kuwajima et al., 1991; Schultz et al., 1992). For RNase HI in the present study, such a phase was not observed, although the native RNase HI has five Pro residues, including one with the *cis* conformation at position 17. Therefore, the isomerization of these Pro residues seems to have only a minimal contribution to the unfolding and refolding processes, resulting in very simple kinetic curves. When there are few parameters describing the folding kinetics, the reliability of the parameters is increased. Therefore, RNase HI is a suitable model protein for folding mechanism analyses.

There are some possible explanations for the fact that the *cis*-proline has little effect on the folding kinetics of RNase HI. (1) Pro17 is in a *cis* conformation, even in the unfolded state. This is not unusual because prolines in several short peptides and the urea-denatured RNase A are largely (more than 70%) in the *cis* conformation, as determined by NMR or by the isomeric specific proteolysis method [reviewed in Brandts and Lin (1986)]. There is evidence that the unfolded states of many proteins have residual structure (Dill & Shortle, 1991; Neri et al., 1992). Therefor, it seems possible that the unfolded state of RNase HI has a residual structure, in which Pro17 is in the *cis* form. (2) The transition in the structure from the intermediate to the native state is largely coupled with proline isomerization. This is equivalent to the assumption that the native or native-like structure with the *trans*-proline does not exist or is very unstable. In this case, a single activation is sufficient for the two transitions, which eliminates the additional phase. Because the activation energy for the proline isomerization (70−90 kJ mol^{-1}; Brandts & Lin, 1986) is comparable to that for the refolding of RNase HI (Figure 8), a large retardation of the process is not expected. Presumably, the tight formation of the

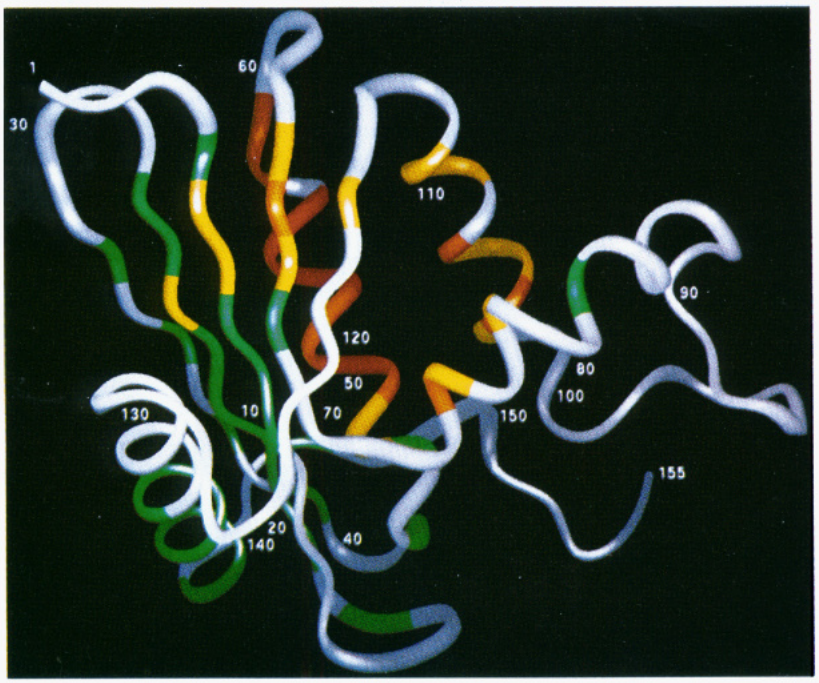

FIGURE 7: Cross-peak intensity ratios after competition at pH 7.5, displayed on the tertiary structure of *E. coli* RNase HI. The residues colored red, orange, yellow, and green have relative cross-peak intensities of more than 0.7, between 0.5 and 0.7, between 0.3 and 0.5, and less than 0.3, respectively. The residues colored white could not be analyzed by NMR, due to the fast H−D exchange in the native state or to the cross-peak overlapping. The crystal structure (Katayanagi et al., 1992) was obtained from the Protein Data Bank (entry 2RN2). The molecular display was accomplished using Insight II (Biosym) on a Silicon Graphics IRIS-4D/25TG computer. The residues colored red are Met50, Ile53, Val54, Ala55, and Leu56. Those colored orange are Leu49, Ala51, Ala52, Ala58, Leu59, Ile66, Val74, Trp104, and Leu107. Those colored yellow are Glu6, Ile7, Ile25, Met47, Glu48, Val65, Leu67, Arg75, Ile78, Leu103, Gln105, Arg106, Ala109, Ala110, Leu111, and Lys117. Those colored green are Val5, Phe8, Thr9, Asp10, Gly18, Gly20, Gly21, Tyr22, Gly23, Ala24, Leu26, Arg27, Lys33, Phe35, Ala37, Tyr39, Thr42, Thr43, Arg46, Ser68, Ile82, Glu135, Leu136, Ala137, Arg138, Ala139, Ala140, Ala141, Met142, and Leu146. α-Helix I (residues 44−59), which is assumed to form in an early stage of the refolding, is the central helix in the molecule and is mainly colored red and orange.

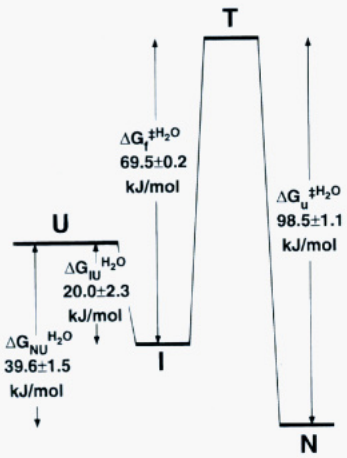

FIGURE 8: Energy profile of unfolding and refolding kinetics under physiological conditions at pH 5.5 and 25 °C. U, I, T, and N represent the unfolded, intermediate, transition, and native states, respectively. $\Delta G_{NU}^{H_2O}$ and $\Delta G_{IU}^{H_2O}$ represent the free energy differences between the native and unfolded states and between the intermediate and unfolded states, respectively. $\Delta G_u^{\ddagger H_2O}$ and $\Delta G_f^{\ddagger H_2O}$ indicate the activation free energies for the unfolding and refolding, respectively.

hydrogen bonds between β-strands A (Glu4−Cys13) and B (Gly20−Tyr29) may be coupled with the *cis*−*trans* isomerization of Pro17.

Formation of α*-Helix I in the Early Refolding Intermediate.* The α-helix I contains a sequence consisting of highly hydrophobic residues, with amide protons that are largely protected from the H−D exchange in the early refolding intermediate. The sequence is Leu49Met50Ala51Ala52-Ile53Val54Ala55Leu56Glu57Ala58Leu59. It has been observed that the N-terminal fragment (Met1−Lys87) has a highly helical structure in aqueous solutions, but the C-terminal fragment (Asp88−Val155) does not (Kanaya & Kanaya, 1995). Furthermore, the peptide Thr34−Lys60, which contains the entire sequence for α-helix I, has a helical structure in 40% acetonitrile at pH 2.0, while the peptide Glu61−Lys86, which contains the sequence for α-helices II and III, does not (Kanaya & Kanaya, 1995). These observations indicate that α-helix I folds independently. Together with the present results, it is very likely that α-helix I is the initiation site of protein folding.

The hydrophobic region in α-helix I is surrounded by hydrophobic residues, such as Ile7, Tyr22, Ala24, Ile26, Phe35, Tyr39, Val65, Leu67, Val103, Trp104, Leu107, Ala110, and Leu111, in the native structure (Katayanagi et al., 1990, 1992; Yang et al., 1990) (Figure 7). These residues are located in β-strands A−D and α-helix IV. Among them, the amide protons of the residues in β-strand D (Val65 and Leu67) and in α-helix IV (Val103, Leu107, and Ala110) are protected more effectively than those of other residues (Figures 6 and 7). Furthermore, the amide protons in α-helices I and IV in the native RNase HI have motions in the millisecond time scale that may be cooperative, as suggested by NMR relaxation analyses (Yamasaki et al., 1995). It is likely that these two helices form a structural unit. Probably, the completion of the initial folding in α-helix I provides a nucleus that facilitates the folding of the surrounding parts, including α-helix IV, through the formation of a hydrophobic core.

Pulsed H−D exchange labeling experiments for RNase T1 (Mullins et al., 1993) and interleukin-1β (Varley et al., 1993) have shown that the amide protons of the residues involved in the hydrophobic core are protected earlier than the other residues. For the acid-induced molten globule of α-lactalbumin, tertiary interactions between the aromatic side chains in the hydrophobic core are observable (Alexandrescu et al., 1993). For the urea-denatured 434 repressor, a hydrophobic core is formed by a slight rearrangement of an α-helix (Neri et al., 1992). Thus, the formation of a hydrophobic core seems to be a general phenomenon observed in the intermediate structure.

The refolding kinetics monitored by the aromatic fluorescence was fit slightly better to the double exponential curve (Figure 1c). This may indicate that the formation of the tertiary structure involves two steps after the formation of the intermediate discussed above. It should be noted that the hydrophobic residues surrounding α-helix I, listed above, include four aromatic residues, Tyr22, Phe35, Tyr39, and Trp104. Especially, Trp104 seems to form a cluster with other Trp residues, such as Trp81, Trp85, and Trp90 (Katayanagi et al., 1990, 1992; Yang et al., 1990). It is possible that the side-chain rearrangements of these aromatic residues require at least two steps, which are reflected in the fluorescence progression.

In a strict sense, there are two possible means of the protection of the amide protons in α-helix I. One is the formation of the hydrogen bonds in the helix. The other is the separation of the amide protons from water by the surrounding hydrophobic residues, without the formation of substantial secondary structure. The former possibility is more likely, because part of the helix (Glu57−Leu59) is not surrounded by hydrophobic residues in the native structure and because the far-UV CD values at 220 nm showed that about 80% of the secondary structure is formed in the intermediate (Figure 1).

It should be noted that the far-UV CD spectra showed that about 80% of the secondary structure is formed (Figure 1a), while the hydrogen-exchange competition method combined with 2D NMR showed that only a restricted region is tightly formed (Figures 6 and 7). It is possible that about 80% of the secondary structure is formed in the intermediate but that the protection of the residues, except those in α-helix I and the surrounding parts, is too weak to be observed by NMR: Suppose that the protection factor (the ratio of the intrinsic H−D exchange rate to that in the protein) of an amide proton is 10, and it becomes replaced with deuterium in as fast as a 500 ms time scale at pH 7.5 and 25 °C (the time scale of the intrinsic H−D exchange rate is 50 ms), while as much as 90% of the proton is involved in a hydrogen bond. It is also possible that hydrogen bonds due to non-native secondary structure are formed in the intermediate, which should be disrupted again during the refolding process. For some proteins, the far-UV CD of the intermediate is more intense than that of the native state (Kuwajima et al., 1987; Radford et al., 1992), for which the formation of non-native α-helices has been suggested (Shiraki et al., 1995).

Energetic and Structural Relationships between the Native, Unfolded, Intermediate, and Transition States. To summarize the equilibrium unfolding and the kinetic folding data, a free energy profile of the refolding and unfolding processes under physiological conditions is illustrated in Figure 8. In this profile, the energy levels of the native, unfolded, intermediate, and transition states are shown. The four energy values indicated in Figure 8 are not independent of each other. Any one of the four values can be calculated using the other three values. There is, however, a discrepancy of 9.4 kJ mol^{-1}, which is above the indicated errors estimated by the fitting error or the noise level. Also, although $m_u^{\ddagger} - m_f^{\ddagger}$ should agree with $m_{NU} - m_{IU}$ ideally, the discrepancy of the m values (6.8 kJ mol^{-1} M^{-1}) exceeds the estimated error. The most probable reason for these discrepancies is another error, such as that generated by the assumption for the extrapolation of the data. In the present study, both the Gibbs free energies of the equilibrium unfolding and the kinetic activation free energy were assumed to have linear dependence on the GuHCl concentration (eqs 4 and 8). If this is not exactly true, it causes a discrepancy in the energy level under physiological conditions. In this case, the m values will have little meaning. For α-lactalbumin, the logarithms of the apparent rate constants have better linearity with the GuHCl concentration than those of the microscopic rate constants (Kuwajima et al., 1989). Furthermore, an H−D exchange study has suggested that, for the dependence of the free energy of equilibrium unfolding on the GuHCl concentration, the curved line proposed by the denaturant binding model (Aune & Tanford, 1969; Nozaki & Tanford, 1970) is more appropriate than a linear line (Bai et al., 1994). Another possible reason for the discrepancy is in the assumed mechanism. In the present study, a three-state mechanism with rapid equilibrium between the unfolded state and one intermediate state is assumed. If there are additional intermediate states, as suggested by the kinetic fluorescence data (Figure 1c), then the mechanism and the microscopic refolding rate constant would change. Therefore, the values obtained here, especially those obtained by extrapolation, should be considered as approximate estimations.

The dependence of the Gibbs free energy difference on the GuHCl concentration (m) correlates with the difference in the accessible surface area (ASA) between the native (or intermediate) and unfolded states (Schellman, 1978). In the present study, m for the transition from the intermediate to the unfolded states is -15.0 ± 1.5 kJ mol^{-1} M^{-1}, while that for the native unfolded transition (Kimura et al., 1992b) is -21.9 ± 0.8 kJ mol^{-1} M^{-1}. This indicates that the ASA of the intermediate is between those of the native and the unfolded states and is closer to that of the native state rather than that of the unfolded state. This is consistent with the molten globule model, which illustrates a compact structure (Ohgushi & Wada, 1983; Kuwajima, 1989).

The ASA of the transition state also can be roughly estimated from the $m^{\ddagger}$ value (Kuwajima et al., 1989). In the present study, the $m^{\ddagger}$ value for the unfolding is -8.0 ± 0.4 kJ mol^{-1} M^{-1}, and that of the refolding is 5.7 ± 0.2 kJ mol^{-1} M^{-1}, which suggests that the ASA of the transition state is roughly the average of those of the intermediate and the native states. By a mutagenic approach, the tertiary interactions in the transition states can be estimated (Matouschek et al., 1989, 1990). This approach may also provide information about the ASA of the transition state of RNase HI.

Comparison of the Intermediate with the Acid-Denatured Molten Globule. It was shown by the CD kinetic progression that the early refolding intermediate is a molten globule (Figure 1). The far-UV CD spectrum (Figure 4) and the

near-UV CD value of the intermediate are similar to that of the acid-denatured molten globule. Furthermore, an H−D exchange experiment has revealed that, in the acid-denatured state, high protection factors (10^2−10^3) are observed only for the residues in α-helix I and the surrounding region (K. Yamasaki, unpublished data). Therefore, the two states have structural similarity. It is clear, however, that they are energetically different: A cooperative unfolding of the intermediate was induced by GuHCl (Figure 3), although the acid-denatured state reveals no cooperative equilibrium unfolding by urea treatment (Dabora & Marqusee, 1994). It is concluded, therefore, that they are similar but not identical. Their structural differences should be revealed, in the future.

Relationships to Mutational Studies. Several mutations have been introduced into α-helix I (Haruki et al., 1994; M. Haruki, unpublished data). Most mutations, such as Leu49→Ile, Leu49→Val, Ala52→Pro, Ala52→Ser, Ala52→Gly, Ala52→Asp, Leu56→Ile, and Leu56→Val, lowered the melting temperature. In contrast, some of the mutations at position 52, such as Ala52→Ile, Ala52→Val, and Ala52→Leu, increased the melting temperature. These results indicate that the hydrophobic interaction in the core region surrounding α-helix I is important for the stability of the native structure. It would be interesting to determine whether these mutations affect the stability of the intermediate against GuHCl, which would yield crucial information about the tertiary structure of the intermediate.

Significance of the Structure of the Intermediate under Physiological Conditions. The denatured state is involved in several biochemical and biological events. For example, it is the preferential target for proteolysis (Imoto et al., 1986), which is necessary to recycle the protein. In general, it seems that the structures in the denatured states are different, depending on the conditions used for the denaturation [reviewed in Dill and Shortle (1991)]. The denaturant-induced unfolded state has less ordered structure than either the acid-induced denatured state or even the thermally denatured state (Aune et al., 1967; Dolgikh et al., 1981; Labhardt, 1982; Kuwajima et al., 1985; Privalov et al., 1989).

In the present study, the early refolding intermediate is shown to be more stable than the denaturant-induced unfolded state, by 20.0 kJ mol^{-1} under physiological conditions (Figure 8). From this ΔG value, the population of the protein in the intermediate state is estimated to be about 3300 times larger than that in the unfolded state, which is equivalent to the denaturant-induced state. Therefore, it is reasonable to assume that the refolding intermediate, rather than the unfolded state, is the latent denatured state under physiological conditions and could be the target for biological processes, such as proteolysis. Structural analyses of the refolding intermediates are therefore significant not only in a biophysical sense but also in a biological sense.

ACKNOWLEDGMENT

The authors thank Prof. K. Kuwajima of the University of Tokyo for helpful advice on the analyses of the folding kinetics and for providing software for the kinetic analyses, Dr. S. Marqusee of the University of California, Berkeley, and Dr. S. Koide of the Scripps Research Institute for helpful discussions and for providing copies of their manuscripts before publication, Dr. T. Ohkubo of our Institute (presently, the Japan Advanced Institute of Science and Technology) for helpful advice on NMR measurements and computations, Drs. M. Haruki, E. Kanaya, and M. Saito of our Institute for helpful discussions, and Mrs. T. Hashimoto-Yamasaki of our Institute for providing portions of the RNase HI protein and some CD spectra.

REFERENCES

Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., & Dobson, C. M. (1993) *Biochemistry 32*, 1707−1718.
Aune, K. C., & Tanford, C. (1969) *Biochemistry 8*, 4586−4590.
Aune, K. C., Salahuddin, A., Zarlengo, M. H., & Tanford, C. (1967) *J. Biol. Chem. 242*, 4486−4489.
Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
Bai, Y., Milne, J., S., Mayne, L., & Englander, S. W. (1994) *Proteins: Struct., Funct., Genet. 20*, 4−14.
Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−199.
Brandts, J. F., & Lin, L.-N. (1986) *Methods Enzymol. 131*, 107−126.
Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
Briggs, M. S., & Roder, H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2017−2021.
Bycroft, M., Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488−490.
Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2nd ed., p 6, W. H. Freeman and Co., New York, NY.
Crouch, R. J., & Dirksen, M. L. (1982) in *Nucleases* (Linn S. M., & Roberts, R. J., Eds.) pp 211−241, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Dabora, J. M., & Marqusee, S. (1994) *Protein Sci. 3*, 1401−1408.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Yu., & Ptitsyn, O. B. (1981) *FEBS Lett. 136*, 311−315.
Eyring, H. (1935) *Chem. Phys. Rev. 17*, 65.
Garel, J.-R., & Baldwin, R. L. (1973) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3347−3351.
Glasoe, P. F., & Long, F. A. (1960) *J. Phys. Chem. 64*, 188−193.
Haruki, M., Noguchi, E., Akasako, A., Oobatake, M., Itaya, M., & Kanaya, S. (1994) *J. Biol. Chem. 269*, 26904−26911.
Ikeguchi, M., Kuwajima, K., Mitani, M., & Sugai, S. (1986) *Biochemistry 25*, 6965−6972.
Imoto, T., Yamada, H., & Ueda, T. (1986) *J. Mol. Biol. 190*, 647−649.
Jennings, P. A., & Wright, P. E. (1993) *Science 262*, 892−896.
Kamath, U., & Shriver, J. W. (1989) *J. Biol. Chem. 264*, 5586.
Kanaya, E., & Kanaya, S. (1995) *J. Biol. Chem. 270*, 19853−19860.
Kanaya, S., & Crouch, R. J. (1983) *J. Biol. Chem. 258*, 1276−1281.
Kanaya, S., & Itaya, M. (1992) *J. Biol. Chem. 267*, 10184−10192.
Kanaya, S., Kimura, S., Katsuda, C., & Ikehara, M. (1990) *Biochem J. 271*, 59−66.
Kanaya, S., Oobatake, M., Nakamura, H., & Ikehara, M. (1993) *J. Biotechnol. 28*, 117−136.
Katayanagi, K., Miyagawa, K., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., & Morikawa, K. (1990) *Nature 347*, 306−309.
Katayanagi, K., Miyagawa, M., Matushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., & Morikawa, K. (1992) *J. Mol. Biol. 223*, 1029−1052.
Kelley, R. F., & Richards, F. M. (1987) *Biochemistry 26*, 6765−6774.
Kiefhaber, T., Quaas, R., Hahn, U., & Schmid, F. X. (1990a) *Biochemistry 29*, 3053−3061.
Kiefhaber, T., Grunert, H.-P., Hahn, U., & Schmid, F. X. (1990b) *Biochemistry 29*, 6475−6480.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.

Kimura, S., Oda, Y., Nakai, T., Katayanagi, K., Kitakuni, E., Nakai, C., Nakamura, H., Ikehara, M., & Kanaya, S. (1992a) *Eur. J. Biochem. 206*, 337−343.
Kimura, S., Nakamura, H., Hashimoto, T., Oobatake, M., & Kanaya, S. (1992b) *J. Biol. Chem. 267*, 21535−21542.
Koide, S., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 12299−12310.
Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87−103.
Kuwajima, K., Hiraoka, Y., Ikeguchi, M., & Sugai, S. (1985) *Biochemistry 24*, 874−881.
Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., & Nagamura, T. (1987) *FEBS Lett. 221*, 115−118.
Kuwajima, K., Mitani, M., & Sugai, S. (1989) *J. Mol. Biol. 206*, 547−561.
Kuwajima, K., Okayama, N., Yamamoto, K., Ishihara, T., & Sugai, S. (1991) *FEBS Lett. 290*, 135−138.
Labhardt, A. M. (1982) *J. Mol. Biol. 157*, 331−355.
Levenberg, K. (1944) *Q. Appl. Math. 2*, 164−168.
Lu, J., & Dahlquist, F. W. (1992) *Biochemistry 31*, 4749−4756.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Marquardt, D. (1963) *SIAM J. Appl. Math. 11*, 431−441.
Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1989) *Nature 340*, 122−126.
Matouschek, A., Kellis, J. T., Jr., Serrano, L., Byroft, M., & Fersht, A. R. (1990) *Nature 346*, 440−445.
Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature 349*, 633−636.
Mullins, L. S., Pace, C. N., & Raushel, F. M. (1993) *Biochemistry 32*, 6152−6156.
Neri, D., Billeter, M., Wider, G., & Wüthrich, K. (1992) *Science 257*, 1559−1563.
Nozaki, Y., & Tanford, C. (1970) *J. Biol. Chem. 245*, 1648−1652.
Ogasahara, K., & Yutani, K. (1994) *J. Mol. Biol. 236*, 1227−1240.
Ohgushi, M., & Wada, A. (1983) *FEBS Lett. 164*, 21−24.
Oobatake, M., Hashimoto, T., Nakamura, H., & Kanaya, S. (1993) *Protein Eng. 6*, 1023.
Pace, C. N. (1990) *Trends Biotechnol. 8*, 93−98.
Palmer, A. G., III, Rance, M., & Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
Privalov, P. L., Tiktopoulo, E. I., Venyaminov, S. Yu., Griko, Y. V., Makhatadze, G. I., & Khechinashvili, N. N. (1989) *J. Mol. Biol. 205*, 737−750.
Ptitsyn, O. B., Pain, R. H., Semisotonov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262*, 20−24.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302−307.
Roder, H., & Wüthrich, K. (1986) *Proteins: Struct., Funct., Genet. 1*, 34−42.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700−704.
Schellman, J. A. (1978) *Biopolymers 17*, 1305−1322.
Schmid, F. X., & Baldwin, R. L. (1979) *J. Mol. Biol. 135*, 199−215.
Schultz, D. A., Schmid, F. X., & Baldwin, R. L. (1992) *Protein Sci. 1*, 917−924.
Shaka, A. J., Barker, P. D., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
Shiraki, K., Nishikawa, & Goto, Y. (1995) *J. Mol. Biol. 245*, 180−194.
Sugawara, T., Kuwajima, K., & Sugai, S. (1991) *Biochemistry 30*, 2698−2706.
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.
Varley, P. V., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H., & Clore, G. M. (1993) *Science 260*, 1110−1113.
Yamasaki, K., Saito, M., Oobatake, M., & Kanaya, S. (1995) *Biochemistry 34*, 6587−6601.
Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H., & Nagayama, K. (1991) *Biochemistry 30*, 6036−6047.
Yamazaki, T., Yoshida, M., & Nagayama, K. (1993) *Biochemistry 32*, 5656−5669.
Yang, W., Hendrickson, W. A., Crouch, R. J., & Satow, Y. (1990) *Science 249*, 1398−1405.
Yutani, K., Ogasahara, K., Suzuki, M., & Sugino, Y. (1979) *J. Biochem. 85*, 915−921.

BI9513721