Structure of the Acid State of *Escherichia coli* Ribonuclease HI†

Jonathan M. Dabora,‡ Jeffrey G. Pelton,§ and Susan Marqusee*,‡

Division of Biochemistry and Molecular Biology, University of California, Berkeley, and Structural Biology Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, Berkeley, California 94720

Received May 16, 1996; Revised Manuscript Received July 24, 1996⊗

ABSTRACT: Under acidic conditions *Escherichia coli* ribonuclease HI* (RNase H*) adopts a partially folded state with all of the properties of a molten globule. Using amide hydrogen exchange carried out under acid state conditions, followed by quenching and NMR detection on the native state, we have determined the residues that are responsible for the observed structure of the acid state. Although RNase H* is a mixed $\alpha + \beta$ protein, a helical subdomain (helices A, D, and B) defines the structure of the acid state. This structure correlates with the rare higher energy conformations detected under native conditions and with data for the earliest intermediates populated in the kinetic folding pathway of the protein.

Molten globules represent partially folded states that are thought to be early intermediates along the folding pathway for most single domain proteins (Ptitsyn et al., 1990; Miranker & Dobson, 1996). In addition to their presumed role in the kinetic pathway of folding, several proteins are known to adopt a molten globule state at equilibrium under extreme solvent conditions such as low pH or by the removal of a ligand or cofactor. All of these molten globules share the following properties: the presence of partial secondary structure and compactness but the lack of fixed side chain interactions that are usually associated with the structures of native proteins (Kuwajima, 1989). The transient nature of kinetic folding intermediates makes them difficult to study. The persistent structure present in equilibrium molten globules, however, represents an attractive model for studying the structural features that govern this important state.

Despite its assumed importance for the structure and folding of a protein, structural data are available for only a small number of molten globule states. To date, indirect studies using amide hydrogen exchange coupled with NMR[1] detection provide the most detailed structural models of molten globule states. Most of these studies have involved either α-helical proteins, such as apomyoglobin (Kuwajima, 1989; Hughson et al., 1990), or proteins with separate helical and sheet domains, such as α-lactalbumin (Chyan et al., 1993). Thus, in order to further characterize both the general and specific properties of molten globule states, we set out to determine the molten globule structure for *Escherichia coli* ribonuclease HI, a protein with a completely different native state motif (mixed $\alpha + \beta$).

E. coli ribonuclease HI (RNase H) is a member of a family of RNases H that cleave RNA from DNA−RNA hybrids in a divalent cation-dependent manner (Hostomsky, 1993). Members of this family are all assumed to adopt a mixed $\alpha + \beta$ motif which has recently been termed the polynucleotidyl transferase fold (Yang & Steitz, 1995). *E. coli* RNase H is a small (155 amino acids), soluble, single domain protein without disulfide bonds. The structure, stability, and dynamics of RNase H are well characterized (Yang et al., 1990; Yamazaki et al., 1991, 1993; Katayanagi et al., 1992; Mandel et al., 1995; Yamasaki et al., 1995b).

In these studies we utilized a variant of the *E. coli* protein (RNase H*) in which all three free cysteines have been replaced by alanine. The cysteines were replaced to minimize any potential complications due to thiol chemistry that can occur during purification and analyses. The activity and structure of RNase H* appear identical to that of the wild-type protein (Kanaya et al., 1990). Previously we identified a partially folded state of RNase H* populated at equilibrium under acid conditions (Dabora & Marqusee, 1994). This acid state of RNase H* has all of the characteristics of a molten globule: it is compact, contains partial secondary structure (∼70% of the native circular dichroism signal), and appears to lack defined tertiary interactions. By altering the conditions, this intermediate can either unfold to the denatured state or refold to its active native state conformation. Furthermore, refolding studies initiated from a urea denatured protein suggest that RNase H* folds through an early molten globule-like intermediate. Refolding studies initiated from the acid intermediate demonstrate rapid folding to the native state, suggesting its potential role as a kinetically competent intermediate (T. M. Raschke and S. Marqusee, personal communication).

To determine the structure of the acid state of RNase H*, we have employed the technique of amide hydrogen exchange on the acid state coupled with indirect NMR detection of amide protons under native conditions. The phenomenon of amide hydrogen exchange has been characterized extensively and has proven to be an effective technique for many different structural studies (Woodward, 1994; Englander et al., 1996). It is particularly powerful for our studies in that it allows structural interpretation with minimal perturbation of the interactions stabilizing the intermediate.

† Supported by a grant from the NIH (GM 50945).
‡ Division of Biochemistry and Molecular Biology.
§ Structural Biological Division.
⊗ Abstract published in *Advance ACS Abstracts,* September 1, 1996.
[1] Abbreviations: NMR, nuclear magnetic resonance; RNase H, ribonuclease HI; RNA, ribonucleic acid; DNA, deoxyribonucleic acid; RNase H*, variant of ribonuclease HI (three free cysteines replaced with alanine); P, protection factor; UV, ultraviolet; CD, circular dichroism; 1D, 2D, and 3D, one dimensional, two dimensional, and three dimensional; HSQC, heteronuclear single-quantum coherence; HMQC, heteronuclear multiple-quantum coherence; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlated spectroscopy; ct-HNCA, constant time HNCA.

S0006-2960(96)01167-1 CCC: $12.00

Amide proton exchange serves as a sensitive probe for the acid state since regions involved in structure, presumably through hydrogen bonding, contain amide hydrogens that are resistant to solvent exchange with deuterium. These amide sites show a measured exchange rate that is slower than expected for the same sequence in a denatured environment. This retardation of exchange is conventionally reported as a protection factor, P, where $P = k_{intrinsic}/k_{observed}$. $k_{intrinsic}$ is the expected hydrogen exchange rate for an unstructured amide (Bai et al., 1993), and $k_{observed}$ is the measured exchange rate. A slower $k_{observed}$ corresponds to a larger protection factor, indicating an increasing amount of structure for a specific amide proton. Hydrogen exchange rates in the acid state were measured by quenching exchange after various times and returning to the native state for determination of proton occupancy by NMR spectroscopy. Only those amide protons which are stable to exchange during the native state NMR detection phase are suitable for determination of acid state protection factors. Using this indirect technique, we find that the molten globule state of RNase H* is comprised of a helical subdomain of the protein. Moreover, this acid state structure is consistent with data on the folding of RNase H and the scarcely populated partially folded structures detected under native conditions.

MATERIALS AND METHODS

Protein Preparation. The plasmid pSM101 was transformed into the T7 expression strain BL21 (DE3), plys S (Novagen). pSM101 directs expression of a variant of *E. coli* ribonuclease HI (RNase H*) in which all three free cysteines have been converted to alanine. Unlabeled protein was produced from cells grown at 37 °C in standard Luria broth with 200 μg/mL ampicillin. Isotopically labeled protein was produced from cells grown at 37 °C in minimum medium (M9) supplemented with 1 mM magnesium sulfate, 100 μM calcium chloride, 0.0010% thiamin, 0.2% glucose, and 200 μg/mL ampicillin (Sambrook et al., 1989). Uniformly ^{15}N-enriched and uniformly ^{15}N/^{13}C-enriched protein samples were obtained using $^{15}NH_4Cl$ (99.8%, 1 g) and $^{15}NH_4Cl$ (99.8%, 1 g)/[$^{13}C_6$]glucose (99.8%, 1 g) per liter of medium, respectively. At a cellular OD of 0.7−0.8 (600 nm) the protein was overexpressed by induction with isopropyl β-thiogalactoside (1 mM) for 3 h. Cells were harvested by centrifugation and purified as described previously (Dabora & Marqusee, 1994) with the following modifications. Heparin−Sepharose fractions containing RNase H* were dialyzed against 50 mM Tris, pH 8.0, 20 mM NaCl, and 0.1 mM EDTA (buffer A) before being loaded onto a Source 15S column (Pharmacia). After being washed with buffer A, RNase H* was eluted using a 200 mL linear gradient from buffer A (20 mM NaCl) to buffer A plus 150 mM NaCl. Fractions containing RNase H* were then concentrated and loaded onto a Sephacryl S-100 gel filtration column (Pharmacia). At this point, the protein was either concentrated by ultrafiltration (Amicon) and stored at −20 °C or lyophilized after dialysis against 50 mM ammonium bicarbonate and stored at 4 °C. Protein concentration was determined by UV absorption spectroscopy (OD_{280} at 1 mg/mL = 2.02) (Kanaya et al., 1990).

Circular Dichroism (CD) Measurements. CD measurements were made on an Aviv 62DS spectropolarimeter equipped with a peltier temperature-controlled sample holder and a 5 mm path-length cuvette. Unfolding of RNase H* was monitored by the ellipticity at 222 nm as a function of pD at 4 °C with samples containing 60 μg/mL RNase H*, 170 mM deuterated phosphoric acid, and 50 mM KCl in D_2O. In all of these studies, pD refers to the isotope-corrected electrode reading. Additional amounts of concentrated deuterium chloride were necessary to reach the pD values for the six most acidic samples (pD 1.4−0.4).

NMR Spectroscopy. All spectra were recorded at 25 °C. Chemical shifts were referenced to H_2O (^{1}H) (4.77 ppm, 25 °C) and indirectly referenced to liquid ammonia (^{15}N) (Live et al., 1984) and 3-([2,2,3,3-$^2H_4$]trimethylsilyl)propionate (TSP) (^{13}C) (Bax & Subramanian, 1986). Quadrature detection in the indirectly detected dimensions was achieved with the States−TPPI method (Marion et al., 1989a), and ^{15}N decoupling was achieved via WALTZ16 modulation of 1.25−1.5 kHz rf fields (Shaka et al., 1983). Data were processed using either the nmrPipe/CAPP suite of programs (Delaglio et al., 1995; Garrett et al., 1991) or Felix (Biosym Technologies).

Two-dimensional ^{15}N−^{1}H correlation (HSQC) spectra (Bax et al., 1990) of RNase H* were recorded with a Bruker AMX-600 spectrometer on uniformly ^{15}N-labeled protein. The H_2O signal was placed on resonance and suppressed using a combination of presaturation and a 1.8 ms spin lock pulse (Messerle et al., 1989). The ^{15}N carrier was set to 120 ppm. For measurement of exchange rates, 32 scans were signal averaged for each complex t_1, t_2 point with a recycle delay of 1.06 s. Each spectrum consists of 512 complex points and a spectral width of 14.1 ppm (^{1}H, t_2) and 128 complex points and a spectral width of 40 ppm (^{15}N, t_1). The total acquisition time for each spectrum was 1.37 h. Spectra were zero-filled to 1K by 1K points and apodized using a 60° shifted sine-bell function. Decay rates for ^{1}H signals using peak intensities (nmrPipe/CAPP) and volumes (Felix) yielded similar results. Peak intensities and volumes were normalized among the various samples for concentration and day to day differences in spectrometer setup by comparing the integrated area of four upfield-shifted methyl resonances observed in 1D spectra acquired immediately after each 2D spectrum.

Three-dimensional ^{15}N-separated TOCSY-HMQC (Driscoll et al., 1990) and NOESY-HMQC (Kay et al., 1989; Marion et al., 1989b) spectra were acquired with a Bruker DMX-600 NMR spectrometer on a 2 mM sample of ^{15}N-labeled RNase H* dissolved in 90% H_2O/10% D_2O (100 mM sodium acetate, pH 5.5). For both experiments, the water was set on resonance and suppressed using presaturation. The ^{15}N carrier was set to 117 ppm. A total of 64 scans were signal averaged for each complex t_1, t_2 point. Each data set consisted of 128 (^{1}H, t_1), 32 (^{15}N, t_2), and 512 (^{1}H, t_3) complex points with spectral widths of 13.5, 27, and 13.5 ppm, respectively. For the TOCSY-HMQC experiment isotropic mixing was achieved via a 100 ms clean-DIPSI sequence (Cavanaugh & Rance, 1992). For the NOESY-HMQC experiment Hα signals that resonate under H_2O were recovered using a 60 ms SCUBA delay (Brown et al., 1988). The NOESY mixing time was set to 100 ms. The data sets were processed using the nmrPipe/CAPP suite of programs. The ^{1}H and ^{15}N dimensions were zero-filled once prior to shifted sine-bell apodization and Fourier transformation. The ^{1}H (t_1) dimension was extended to 256 points using complex linear prediction prior to processing.

A 3D constant-time HNCA spectrum (Grzesiek & Bax, 1992) was recorded as described on an AMX-600 spectrometer equipped with a three-channel interface, external 50 W linear amplifier (ENI, Rochester, NY) for ^{13}C frequencies, and a triple-resonance ^{13}C optimized probe. ^{1}H, ^{15}N, $^{13}C\alpha$, and ^{13}CO frequencies were set to H_2O (4.77 ppm), 117, 56, and 177 ppm, respectively. ^{1}H 180° pulses were used for proton decoupling. The data matrix consisted of 512 (t_3, ^{1}H), 22 (^{15}N, t_2), and 52 (^{13}C, t_1) complex points with spectral widths of 14.1, 27.0, and 30.0 ppm, respectively. Carbonyl pulses were applied as off-resonance-shaped pulses with an excitation profile corresponding to the first lobe of a sinc function. The 3D data set was processed with the nmrPipe suite of programs. Data were zero-filled once in the ^{1}H and ^{13}C dimensions, and mirror image linear prediction (Zhu & Bax, 1990) was used to extend the ^{15}N dimension from 22 to 40 complex points prior to zero-filling (64 complex points), apodization, and Fourier transformation. The resulting matrix consisted of 1024 (^{1}H), 64 (^{15}N), and 128 (^{13}C) real points.

Native State Hydrogen Exchange. Native state hydrogen exchange rates for the amide protons of RNase H* were determined on a single sample of uniformly ^{15}N-labeled protein. RNase H* was lyophilized in the folded state from H_2O (50 mM ammonium bicarbonate) and resuspended to a final concentration of 10 mg/mL RNase H* and 100 mM NaOAc in D_2O at pD 5.5. 2D $^{15}N-^{1}H$ HSQC spectra were acquired at 11 different time points after dissolution in deuterated buffer at 25 °C (86, 208, 381, 698, 1278, 2853, 8663, 12 907, 38 898, 69 510, and 187 225 min). The aforementioned time points are defined as the midpoint of the acquisition of each HSQC experiment. Amide proton exchange rates were calculated by fitting the decay in proton occupancy to a single exponential function. Both peak volume and peak height were used to calculate proton occupancy without much change in the results. Only 39 slowly exchanging amide protons, those with a half-life >5.3 days (or <2% change in 24 h), were used as probes for the acid state exchange.

Acid State Hydrogen Exchange. Amide hydrogen exchange in the acid state was initiated by resuspending 50 mg of lyophilized, protonated, ^{15}N-labeled RNase H* to a final concentration of 3 mg/mL in 170 mM deuterated phosphoric acid, 50 mM KCl, and 45.5 mM DCl in D_2O (pD = 1.26) at 4 °C. After 1.38, 1.92, 2.92, 6, 21, 61, 160, 600, and 1560 min, hydrogen exchange was quenched by removing 5 mg aliquots and diluting to native conditions (0.4 mg/mL RNase H* in 100 mM sodium acetate in D_2O). Each sample was concentrated to ~10 mg/mL by ultrafiltration via Centriprep 10 and Centricon 10 concentrators (Amicon), and the pD of each sample was adjusted to pD 5.5. In order to minimize the effects of native state exchange, NMR spectra were acquired within 24 h. All experimental manipulations were carried out at 4 °C, and samples were kept at 4 °C prior to NMR detection. Exchange rates for each amide hydrogen probe were determined as described for the native state. The estimated error for these rates is ~2-fold based on a comparison of proton occupancy determined by peak height versus peak volume. Protection factors were determined from the measured $k_{observed}$ by following the equation: $P = k_{observed}/k_{intrinsic}$. The calculated $k_{intrinsic}$ values correspond to the unstructured exchange rates which correct for pH, temperature, and nearest neighbor side chain exchange effects (Bai et al., 1993).

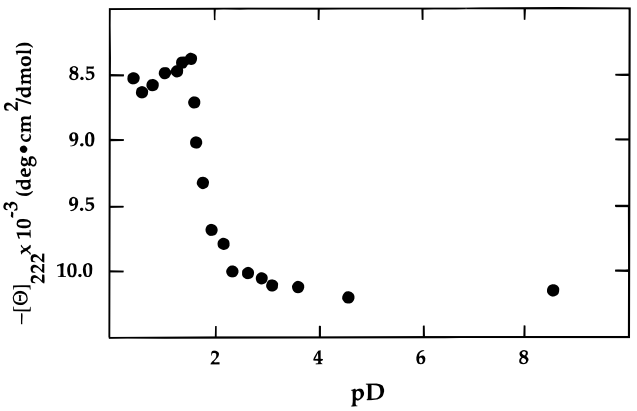

FIGURE 1: pD dependence of the CD signal at 222 nm for RNase H* (170 mM deuterated phosphoric acid, 50 mM KCl, 60 μg/mL protein in D_2O at 4 °C).

RESULTS

Acid State of RNase H at Low pD. A pD titration of RNase H* in deuterated solvent revealed a partially folded state as monitored by the circular dichroism signal at 222 nm (Figure 1). This result is consistent with similar experiments that first identified a molten globule state for RNase H* under acidic conditions in protic buffer (Dabora & Marqusee, 1994). Stabilization of the acid state occurred with a transition midpoint of pH ~1.8 in either deuterated or protic conditions. The partial structure observed in the deuterated acid state underwent reversible thermal denaturation (data not shown). Therefore, under these conditions RNase H* has the same molten globule characteristics observed previously in protic conditions. The buffer conditions used for this pD titration are identical to those used for the acid state hydrogen exchange reactions described below.

NMR Assignments. NMR assignments of the amide protons in the native state of RNase H* were needed to provide specific structural probes for the hydrogen exchange studies. The three cysteine to alanine substitutions in RNase H* resulted in a perturbation in the NMR spectra; therefore, the previous assignments of wild-type RNase H did not correlate directly with the peaks of RNase H* (Yamazaki et al., 1993). Thus, backbone NH, CαH, and Hα resonances for RNase H* were assigned via 3D NOESY, 3D TOCSY, and 3D HNCA experiments. This strategy allowed the identification of 138 of the 150 backbone amide hydrogens (>92%). Backbone amide hydrogen chemical shift differences between RNase H* and wild-type RNase H of greater than 0.1 ppm were observed for 28 of the assigned backbone amide protons with 3 amide protons having chemical shift differences greater than 0.3 ppm. These assignments are available through the BioMagResBank database (accession number: bmr4012).

Approximately half of the $^{15}N-^{1}H$ cross-peaks were assigned by direct comparison with the wild-type assignments. These assignments were confirmed and extended using ^{15}N-edited 3D NOESY, ^{15}N-edited 3D TOCSY, and ^{15}N-edited 3D HNCA spectra. The intraresidue and interresidue NOE correlations used to make sequential assign-

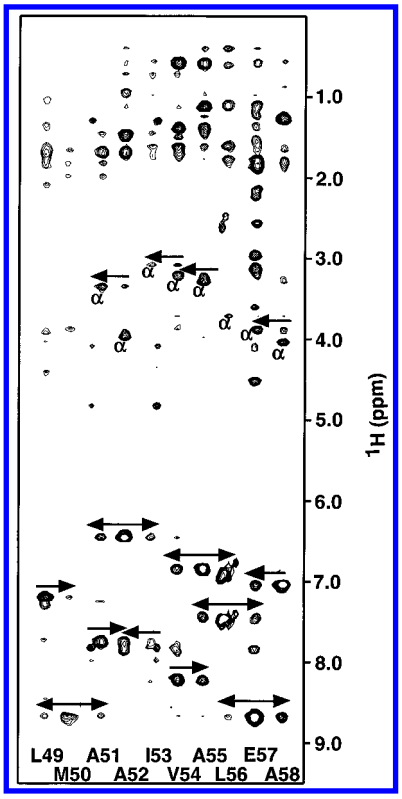

FIGURE 2: Strip plot of a portion (helix A) of a 3D ^{15}N-separated NOESY-HMQC spectrum of RNase H*. Arrows indicate d_{NN} and $d_{\alpha N}$ NOEs used to obtain sequential assignments. Intraresidue HN−Hα NOEs are denoted by an α. ^{15}N−^{1}H assignments are listed at the bottom of the figure.

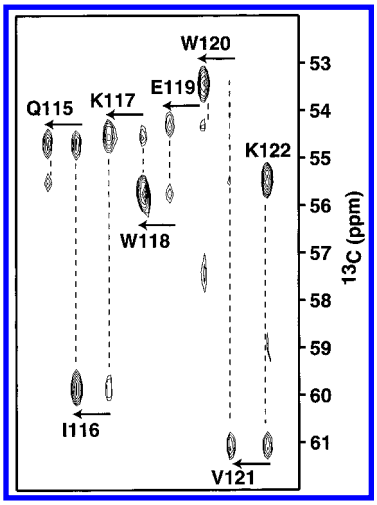

FIGURE 3: Strip plot of a portion (β-strand 5) of a 3D ct-HNCA spectrum of RNase H* used to determine sequential assignments. Intraresidue HN−Cα correlations are denoted by one-letter amino acid code and residue number. Arrows connect interresidue HN−Cα(i−1) correlations with the intraresidue correlation of the preceding residue. Note that the interresidue correlation for V121−W120 is missing. It is observed at lower contours.

ments for residues L49−A58 in helix A are shown in Figure 2. For residues Q115−K122 in strand 5 of the β-sheet region, both intraresidue and interresidue HN−Cα connectivities from the 3D HNCA spectrum were used to confirm and extend HN assignments determined from 3D NOESY and 3D TOCSY spectra (Figure 3).

The secondary structure patterns identified by the NOESY experiment on RNase H* are consistent with previously published experiments on wild-type RNase H (Yamazaki et al., 1991). These regions of secondary structure are also consistent with the mixed α + β motif determined by X-ray crystallography on the wild-type protein (Katayanagi et al., 1992). Therefore, the high-resolution crystal structure of RNase H provides a reasonable basis for interpreting our hydrogen exchange studies. This similarity between RNase H* and wild-type RNase H was expected since the three cysteine to alanine mutations do not appear to affect the activity, CD spectrum, or tryptophan fluorescence of the protein (Kanaya et al., 1990).

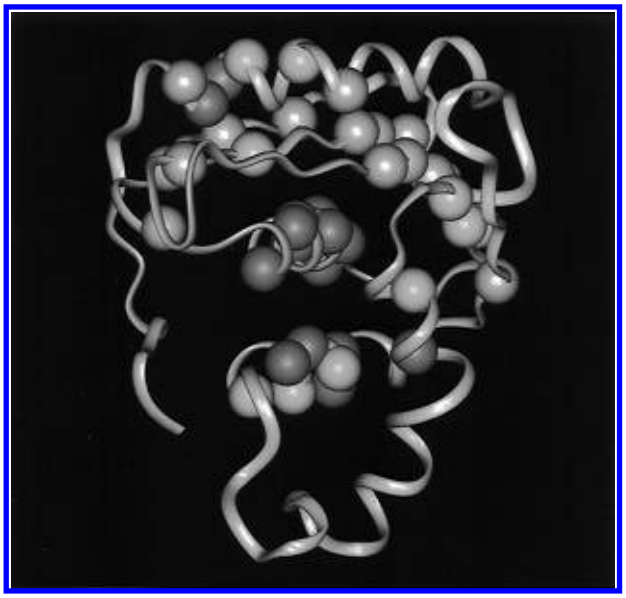

FIGURE 4: Protection factors of acid state amide probes modeled onto the native crystal structure of ribonuclease HI (Katayanagi et al., 1992). Amide hydrogen probes for the acid state structure are represented as spheres on the backbone ribbon. The most structured regions of the acid state (P > 10) are highlighted in red. Amide protons that reside in unstructured regions of the acid state (P < 5) are shown in gray. Amide protons with intermediate levels of protection (5 < P < 10) are shown in blue.

Twelve amide protons of RNase H* could not be assigned. Six of these protons reside at the N-terminus; therefore, it is likely that the N-terminus of the protein is unstructured in solution. The other missing assignments are scattered throughout the protein sequence (D10, T43, A63, E131−A133).

Identification of Slowly Exchanging Amide Hydrogens in the Native State. The slowly exchanging protons identified by amide hydrogen exchange on the native state of RNase H* (pD 5.5) were used as probes for our indirect study of the acid state structure. A 2D ^{15}N−^{1}H HSQC spectrum of the native state of RNase H* provided sufficient dispersion to measure these exchange rates. NMR spectra were acquired at 11 different time points between 1.5 h and 130 days (see Materials and Methods). Forty-five amide protons exchanged slowly enough in the native state to serve as acid state probes: they all have a half-life greater than 5.3 days and a corresponding protection factor $P > 5 \times 10^4$. Of these 45 protected protons, only 39 were used as probes due to problems with either spectral overlap or confidence in the exchange measurements.

All of the slowly exchanging protons, or probes, reside within elements of secondary structure. These 39 probes are scattered throughout the protein with ∼40% in β-sheet regions and ∼60% in α-helical regions (Figure 4). The only element of secondary structure that lacks a probe is the eight-

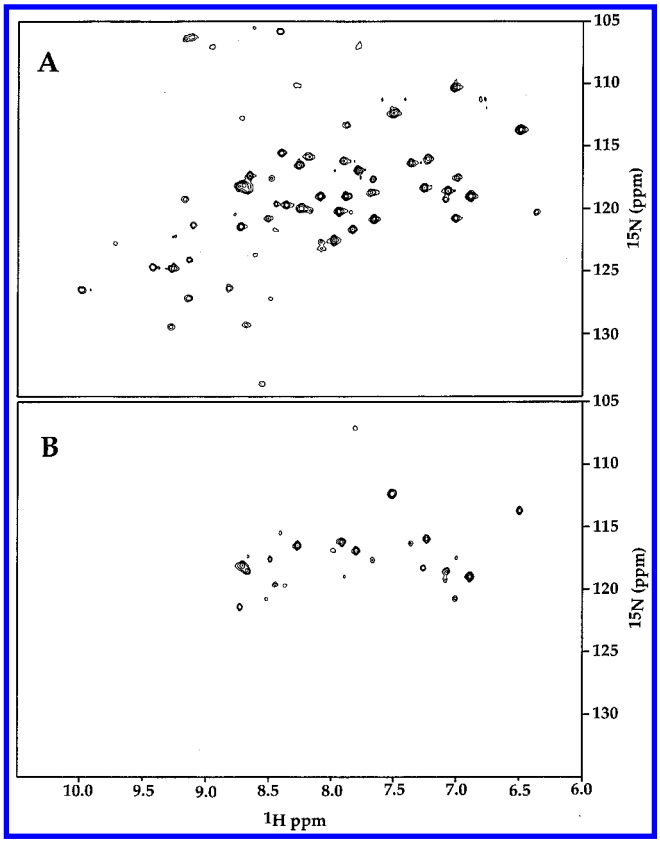

FIGURE 5: $^{15}N-^{1}H$ HSQC spectra acquired 1.38 min (A) and 160 min (B) after initiation of hydrogen exchange from the acid state.

residue helix C. This well-dispersed distribution of probes allowed us to carry out a detailed structural analysis of the molten globule state.

Determination of Amide Hydrogen Exchange Rates in the Acid State. In order to determine hydrogen exchange rates in the acid state, exchange was carried out under molten globule conditions (pD 1.3) and then quenched by raising the pD to native conditions (pD 5.5). Quenching of exchange provides both the necessary dispersion of amide protons in the native state and the ability to evaluate short time points, which is important for determining rates of rapidly exchanging protons. As stated above, only those protons that do not exchange while the protein is in the native state can serve as suitable probes for these acid state exchange measurements. Aliquots were removed from the acid state exchange reaction at nine different time points ranging from 1 min to 24 h (see Materials and Methods). The loss of protons over time is easily seen by comparing the HSQC spectra for two different time points (1.38 and 160 min) in Figure 5. While some amide hydrogen peaks are completely absent in the later time point, other proton peaks show a measurable decrease in proton occupancy, and a few amide hydrogens are almost completely resistant to hydrogen exchange and show no change in peak intensity. The exchange data for each acid state probe were fit to a single exponential to determine $k_{observed}$, and the calculated acid state protection factors are summarized in both Table 1 and Figure 6.

Residues in helix A comprise the most protected region of the acid state of RNase H*. All eight amide hydrogen probes in helix A (residues R46−E57) are significantly resistant to hydrogen exchange with protection factors greater than 45. Helix D, which in the native state pairs with helix A in a coiled-coil motif, is also structured in the acid state.

Table 1: Hydrogen Exchange Protection Factors, *P*, for the Acid State of RNase H*

residue	*P*	residue position in native state	residue	*P*	residue position in native state
Ile 7	4	strand 1	Ile 66	8	strand 4
Phe 8	2	strand 1	Leu 67	7	strand 4
Thr 9	1	strand 1	Ser 68	6	strand 4
Gly 20	3	strand 2	Arg 75	7	helix B
Tyr 22	7	strand 2	Ile 78	29	helix B
Ile25	2	strand 2	Val 101	7	helix D
Leu 26	2	strand 2	Trp 104	21	helix D
Arg 27	2	strand 2	Gln 105	7	helix D
Phe 35	3	strand 3	Leu 107	12	helix D
Ala 37	≤2	strand 3	Asp 108	8	helix D
Tyr 39	≤2	strand 3	Ala 109	35	helix D
Thr 42	2	loop	Ala 110	19	helix D
Arg 46	45	helix A	Lys 117	4	strand 5
Glu 48	45	helix A	Asp 134	2	helix E
Ala 51	196	helix A	Leu 136	7	helix E
Ala 52	183	helix A	Ala 139	8	helix E
Val 54	87	helix A	Ala 140	14	helix E
Ala 55	252	helix A	Ala 141	15	helix E
Leu 56	343	helix A	Met 142	4	helix E
Glu 57	121	helix A			

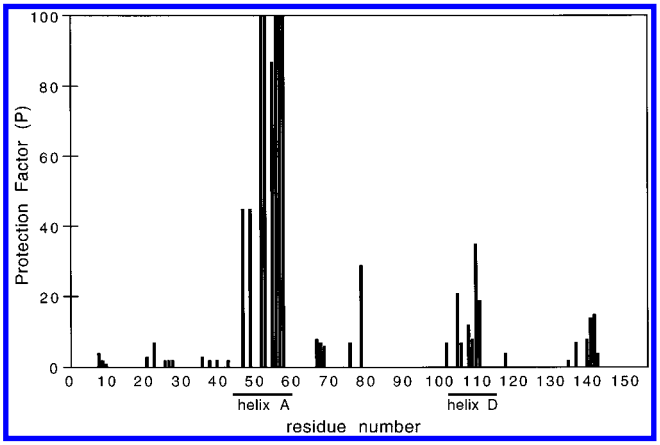

FIGURE 6: Amide hydrogen exchange protection factors for the acid state of ribonuclease HI.

Moderate to high levels of protection ($7 > P > 35$) are observed for all seven amide protons in helix D. Amide protons for R75 and I78 in helix B have moderate to low levels of protection. Helix E, which is at the C-terminus of RNase H shows four out of six amide protons with moderate to low levels of protection. This helix forms no direct contact with the other helices in the native state. No information is available about the structural contribution of helix C due to the absence of unambiguous amide proton probes.

Equally significant to these protected regions is the absence of protected amide protons ($P \leq 5$) in other regions of the protein. For example, the majority of amide probes in the β-sheet region of the protein show no protection. Therefore, the β-sheet of RNase H does not contribute to the acid state structure. Interestingly, it is the β-sheet that separates helix E from the well-protected regions of the acid state. This suggests that the low levels of protection in helix E may well be due to its intrinsic helical propensity.

In summary, a structured subdomain of RNase H* which involves residues in helices A, B, and D appears to stabilize the acid state structure. In addition, the β-sheet region of the protein does not appear to contribute to the structure of this state. Figure 4 displays a model of the acid state in the context of the native state structure.

DISCUSSION

Under acidic conditions, *E. coli* RNase H* adopts a partially folded state with all the characteristics of a molten globule: partial secondary structure, lack of direct tertiary interactions, and compactness (Dabora & Marqusee, 1994). Using amide hydrogen exchange techniques coupled with multidimensional NMR detection, we have indirectly determined a structural model for this partially folded acid state. Our results suggest that, as seen for several other molten globules, only a select region or subdomain of the native structure is responsible for the stability and structure of the acid state of RNase H* (Figure 4). This subdomain is limited to an α-helical subregion of this mixed $\alpha + \beta$ protein.

*Structured Regions in the Acid State of RNase H**. Structure in the acid state of RNase H* was determined by performing hydrogen exchange under acid state conditions and then quenching to the native state for NMR detection of amide proton occupancy. Since some amides exchange easily under native state conditions, information about the acid state structure is limited to a subset of the amide protons that do not significantly exchange during the detection phase of the experiment. These 39 amide protons are distributed throughout the native structure of the protein (see Figure 4).

The acid state of RNase H* represents an ensemble of structures distinct from the native state. Amide protons within helices A, B, D, and E show varying degrees of protection in the acid state, while amide protons in the β-sheet show an almost complete lack of protection (Figure 6). These protection factors cannot be attributed to a global destabilization of the native state as the pattern of exchange in the acid state is significantly different from that in the native state (Chamberlain et al., 1996). In fact, some of the most protected sites in the native state of the protein reside in the β-sheet.

The most structured region of the acid state is clearly helix A. All eight amide probes in helix A are the most highly protected amide protons in the acid state, and all have protection factors greater than 45. A subset of amide hydrogens, A51, A52, A55, L56, and E57, have protection factors greater than 100. Protection factors of this magnitude are considered high for a molten globule state, and they provide evidence that helix A is significantly structured. Protection at the N-terminus of helix A decreases, suggesting a potential lack of stability or fraying at that end. Helix D comprises the next most stable region of the acid state with many amides showing protection factors from 10 to 50. These moderate protection factors are similar in size to those previously observed for structured regions in molten globules such as apomyoglobin (Hughson et al., 1990) and α-lactalbumin (Chyan et al., 1993; Schulman et al., 1995). Helix E contains acid state amide probes that have low to moderate levels of protection (2−15).

Most of the amide protons in the β-sheet region of RNase H* show no detectable protection in the acid state ($P < 5$), and no residues in the sheet have a protection factor greater than 10. Therefore, the β-sheet region does not appear to contribute to the structure and stability of this state. Taken together with data for the α-helices, these observed protection factors suggest that the molten globule of RNase H* is comprised of a marginally stable helical subdomain of the native structure and allow us to create a structural model for this acid state.

Structural Model for the Acid State. The pattern of protection factors in the acid state suggests that the structure of the molten globule is stabilized through an apparent interaction of hydrophobic residues present in a subset of helical elements of the protein. Helix A, which is at the center of the native state hydrophobic core, appears to be the most crucial element for stability in the acid state. Since such significant protection levels have never been observed in isolated helices, these residues must be stabilized by other regions of the protein. In the native state of RNase H*, this helix packs against helix D to form a coiled-coil motif (O'Shea et al., 1991). Helix D is the region with the second highest degree of protection in the acid state. It seems reasonable to assume that these two helices are also associated with each other in the acid state. Given the uncertainty in our determined rates, we are unable to discriminate if the pattern of protection suggests native-like packing interactions, or not. Association of these two helices in the acid state, which also form the hydrophobic core of the native state, however, suggests that both intrinsic helical propensity and hydrophobicity serve to stabilize the molten globule. The two acid state probes for helix B, which also interacts with helices A and D in the native state (Figure 4), have low to moderate levels of protection, suggesting that helix B also contributes to the structure of the acid state.

While these hydrogen exchange experiments cannot provide information about potential structure in helix C and the adjacent loop, fluorescence data imply that residues in this region may play a role in the acid state structure. RNase H* contains six tryptophans. Four of the tryptophans, two from helix C, one from the adjacent loop, and one from helix D, form an interacting cluster in the native structure (Katayanagi et al., 1992). Since residual tryptophan fluorescence detected in the acid state is lost by the addition of chemical denaturants (Dabora & Marqusee, 1994), and the remaining two tryptophan residues reside in the unstructured region of strand 5, it seems reasonable to suggest that the residual tryptophan fluorescence present in the acid state arises from these four interacting tryptophans. This cluster, however, is not necessary for formation of a stable acid state since a mutant that lacks residues 83−95 (parts of helix C and the adjacent loop) can still form an acid state molten globule (unpublished results).

Several residues in helix E show low levels of protection and therefore may contribute structure to the acid state. In the native structure of RNase H*, however, helix E is separated from the core of the acid state structure (helices A and D) by the β-sheet which shows no protection. Residues in helix E show a high helical potential, and indeed helix E has been shown to form a stable helix when isolated as a peptide in both acid and neutral conditions (Eric Goedken and S. Marqusee, personal communication). This suggests that the low level of protection seen for helix E may arise from its intrinsically high helical propensity. Furthermore, the high level of protection seen in helix A and helix D underscores the importance of a hydrophobic core for the stability of the molten globule. In conclusion, our model of the acid state suggests that helices A and D directly interact with one another to stabilize a subdomain with some contribution from nearby helix B. From the data

alone, however, it cannot be ruled out that helix E makes nonnative interactions with the other protected regions of the acid state.

Relationship to Other Acid States. Our model for RNase H* shares many general characteristics with the few other structural models for nonnative acid states. For all of these proteins, the molten globule state appears to comprise a subdomain of the native state. For instance, in the molten globule state of apomyoglobin, helices A, G, H, and, under some conditions, B define the molten globule state (Hughson et al., 1990; Loh et al., 1995). In α-lactalbumin (Chyan et al., 1993) and equine lysozyme (Morozova et al., 1995), both α/β proteins, the structure in the molten globule appears to be limited to the α-domain. Although the native state of RNase H contains a mixed α + β fold, the acid state structure seems to also be limited to a helical subset of the protein. These results lead one to suggest that local intrinsic helical tendencies play a major driving force in stabilizing molten globules, and more specific tertiary packing interactions are needed to stabilize a β-sheet.

The Acid State Molten Globule Exists as a High-Energy State under Native Conditions. Recently, by examining the rates of amide hydrogen exchange as a function of low denaturant concentration, we have identified the most stable regions of RNase H* under native conditions (pH 5.5) (Chamberlain et al., 1996). By sequentially folding these regions, we see that helices A and D define a partially folded state with a marginal stability of 1.3 kcal/mol greater than the unfolded state. Helix B and strand 4 define the next most stable region of the protein and comprise a partially folded state consisting of helices A, B, D, and strand 4 with a stability of 2.6 kcal/mol. The completely folded protein or native structure has a stability of ∼10 kcal/mol under these conditions. These partially folded species under native conditions, while rarely populated, correlate remarkably well with the pattern of amide protection that we have determined for the molten globule state, which is well populated under acid conditions. Helices A and D form the most structured region in the acid state ensemble with helix B showing a somewhat less level of protection. While none of the β-sheet shows significant levels of protection, the three probes in strand 4 are the most protected in the β-sheet ($P = 6-8$). Hence, the acid state molten globule, or an ensemble with similar protection patterns, represents the high-energy conformations populated under native conditions. The physical basis for the relative destabilization of the native structure upon the addition of protons remains to be uncovered.

The Acid State as a Kinetic Folding Intermediate. Equilibrium molten globules are often associated with general kinetic intermediates populated early in the folding pathway. This hypothesis has been clearly demonstrated for only one protein, apomyoglobin. Within 5 ms of initiating folding, RNase H* folds to an early molten globule like intermediate. This marginally stable intermediate shows ∼70% of the native CD signal and binds ANS (Tanya Raschke and S. Marqusee, unpublished results). Competition hydrogen exchange studies on the folding of wild-type ribonuclease HI (with three free cysteines) demonstrate that, at pH 7.9, helix A is the first to acquire protection, followed by helices D, B, and strand 4 (Yamasaki et al., 1995a). These studies suggest that for RNase H the early kinetic intermediates resemble the acid state molten globule.

In conclusion, our hydrogen exchange studies suggest that the structure in the acid molten globule of RNase H* is limited to a subdomain of the native structure. Although RNase H* is a mixed α + β protein, this subdomain is comprised of only an α-helical portion of the protein. We cannot determine if this subdomain has native-like topology, although CD and ANS fluorescence data imply that it lacks the well-defined tertiary interactions present in the native state. The correlation we see between the high-energy conformations weakly populated under native conditions, kinetic folding studies, and the acid state molten globule supports the widely held notion that equilibrium molten globules serve as good model systems for protein folding intermediates. Future experiments using site-specific mutations in RNase H* can address the role of the molten globule in protein folding.

ACKNOWLEDGMENT

We thank Dave Wemmer for support and use of NMR time, Tracy Handel for help with NMR, and Aaron Chamberlain for help with analysis of exchange data. We also thank Tanya Raschke for critical reading of the manuscript and all the members of the Marqusee laboratory for stimulating discussion.

REFERENCES

Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins 17*, 75−86.
Bax, A., & Subramanian, S. (1986) *J. Magn. Reson. 67*, 565−569.
Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304−318.
Brown, S. C., Weber, P. C., & Mueller, L. (1988) *J. Magn. Reson. 77*, 166−169.
Cavanaugh, J., & Rance, M. (1992) *J. Magn. Reson. 96*, 670−678.
Chamberlain, A. C., Handel, T. M., & Marqusee, S. (1996) *Nat. Struct. Biol.* (in press).
Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A., & Baum, J. (1993) *Biochemistry 32*, 5681−5691.
Dabora, J. M., & Marqusee, S. (1994) *Protein Sci. 3*, 1401−1408.
Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 3542−3556.
Englander, S. W., Sosnick, T. R., Englander, J. J., & Mayne, L. (1996) *Curr. Opin. Struct. Biol. 6*, 18−23.
Garrett, D. S., Powers, R., Gronenborn, A. M., & Clore, G. M. (1991) *J. Magn. Reson. 95*, 214−220.
Grzesiek, S., & Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
Hostomsky, Z., Hostomska, Z., & Matthews, D. A. (1993) *Ribonucleases H*, Cold Spring Harbor Laboratory Press, Plainview, NY.
Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544−1548.
Jennings, P. A., & Wright, P. E. (1993) *Science 262*, 892−896.
Kanaya, S., Kimura, S., Katsuda, C., & Ikehara, M. (1990) *Biochem. J. 271*, 59−66.
Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., & Morikawa, K. (1992) *J. Mol. Biol. 223*, 1029−1052.
Kay, L. E., Marion, D., & Bax, A. (1989) *J. Magn. Reson. 84*, 72−84.
Kuwajima, K. (1989) *Proteins 6*, 87−103.
Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1934−1941.
Loh, S. N., Kay, M. S., & Baldwin, R. L. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5446−5450.

Mandel, A. M., Akke, M., & Palmer, A. G. I. (1995) *J. Mol. Biol. 246*, 144−163.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989a) *J. Magn. Reson. 85*, 393−399.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.
Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608−613.
Miranker, A. D., & Dobson, C. M. (1996) *Curr. Opin. Struct. Biol. 6*, 31−42.
Morozova, L. A., Haynie, D. T., Arico, M. C., Van, D. H., & Dobson, C. M. (1995) *Nat. Struct. Biol. 2*, 871−875.
O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. (1991) *Science 254*, 539−544.
Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262*, 20−24.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schulman, B. A., Redfield, C., Peng, Z. Y., Dobson, C. M., & Kim, P. S. (1995) *J. Mol. Biol. 253*, 651−657.
Shaka, A. J., Keeler, J., & Freeman, R. (1983) *J. Magn. Reson. 53*, 313−340.
Woodward, C. K. (1994) *Curr. Opin. Struct. Biol. 4*, 112−116.
Yamasaki, K., Ogasahara, K., Yutani, K., Oobatake, M., & Kanaya, S. (1995) *Biochemistry 34*, 16552−16562.
Yamasaki, K., Saito, M., Oobatake, M., & Kanaya, S. (1995b) *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., & Steitz, T. A. (1995) *Structure 3*, 131−134.
Yang, W., Hendrickson, W. A., Crouch, R. J., & Satow, Y. (1990) *Science 249*, 1398−1405.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405−410.

BI9611671