Equilibrium unfolding of *Escherichia coli* ribonuclease H: Characterization of a partially folded state

JONATHAN M. DABORA AND SUSAN MARQUESEE

Department of Molecular and Cell Biology, Division of Biochemistry and Molecular Biology,
University of California–Berkeley, Berkeley, California 94720

(RECEIVED March 9, 1994; ACCEPTED June 21, 1994)

Abstract

We have examined the equilibrium unfolding of *Escherichia coli* ribonuclease HI (RNase H), a member of a family of enzymes that cleaves RNA from RNA:DNA hybrids. A completely synthetic gene was constructed that expresses a variant of the wild-type sequence with all 3 cysteines replaced with alanine. The resulting recombinant protein is active and folds reversibly. Denaturation studies monitored by circular dichroism and tryptophan fluorescence yield coincident curves that suggest the equilibrium unfolding reaction is a 2-state process. Acid denaturation, however, reveals a cooperative transition at ~pH 1.8 to a partially folded state. This acid state can be further denatured in a reversible manner by the addition of heat or urea as monitored by either CD or tryptophan fluorescence. Analytical ultracentrifugation studies indicate that the acid state of RNase H is both compact and monomeric. Although compact, the acid state does not resemble the native protein: the acid state displays a near-UV CD spectrum similar to the unfolded state and binds to and enhances the fluorescence of the dye 1-anilinonaphthalene, 8-sulfonate much more than either the native or unfolded states. Therefore, the acid state of *E. coli* RNase H has the characteristics of a molten globule: it retains a high degree of secondary structure, remains compact, yet does not appear to contain a tightly packed core.

Keywords: acid state; molten globule; protein folding; ribonuclease H

The native conformation of a protein arises from many interdependent coupled interactions. A number of forces act in a cooperative manner to increase the stability of the folded state under native conditions. Therefore, in order to understand protein stability fully, we must understand which interactions are most important. This has proven to be difficult because the energetic importance of any particular residue is dependent on its sequence and structural context. Similarly, the cooperativity of the folding process also makes the study of transient intermediates difficult. One approach to circumvent these problems imposed by cooperativity is to study equilibrium conformations of the molecule that are only partially folded. Such intermediate states will yield important information on the range of contexts in which the same amino acid sequence will fold into a stable structure and should reveal which interactions are driving the folding process. These equilibrium intermediates may also serve as good experimental models for transiently populated intermediates in the kinetic pathway of folding.

The molten globule is 1 such example of a partially folded equilibrium state. This non-native state of a protein is defined as having an almost native level of secondary structure and compactness, yet lacks the well-packed core characteristic of native proteins. The molten globule state is also thought to be a general intermediate formed early in the kinetic pathway of folding. In this study, we have initiated a search for such non-native states of the *Escherichia coli* protein, ribonuclease HI (RNase H).

RNase H is an endoribonuclease that specifically degrades the RNA portion of DNA:RNA hybrids (Crouch & Dirksen, 1982). The *E. coli* protein is a member of a large family of enzymes with high sequence homology in both prokaryotes and eukaryotes (Johnson et al., 1986) and appears to be an excellent model system for protein folding studies. It is a small (155 amino acids) single-domain protein that contains no disulfide bonds. High-resolution structural data are available from both X-ray crystallography, now at 1.48 Å resolution (Yang et al., 1990; Katayanagi et al., 1992), and NMR studies (Yamazaki et al., 1991). These structural studies reveal a single-domain $\alpha + \beta$ protein not homologous to other known structures outside of the RNase H family (Fig. 1). The thermodynamic stability of the wild-type *E. coli* protein has been determined. In addition, mutations that increase and decrease protein stability have been ob-

Reprint requests to: Susan Marqusee, Department of Molecular and Cell Biology, Division of Biochemistry and Molecular Biology, 229 Stanley Hall, University of California–Berkeley, Berkeley, California 94720; e-mail: marqusee@ucxray6.berkeley.edu.

Fig. 1. Ribbon diagram of the crystal structure of *E. coli* RNase H (drawn with MOLSCRIPT [Kraulis, 1991]).

tained and characterized (Kanaya et al., 1991; Ishikawa et al., 1993).

Although the physiological function of the bacterial RNase H is unclear, in retroviruses the activity resides as a domain within the reverse transcriptase molecule and is essential for the retroviral life cycle. The crystal structure of the domain from HIV shows the same basic topology as the *E. coli* protein (Davies et al., 1991), and, by analogy, we assume that all of the RNase H domains contain the same structural motifs.

In this study, we present a thermodynamic characterization of RNase H in an effort to identify populated equilibrium intermediates. The protein used here is a variant of the *E. coli* sequence in which all 3 free cysteines have been replaced by alanine. These 3 cysteines are solvent-exposed, and Kanaya et al. (1990) have shown that changing them to alanine has no effect on the activity or the CD spectrum. Our studies reveal that this RNase H forms a partially folded state at low pH, a so-called "acid state." We have characterized this acid state of RNase H and show that it has all of the properties of the recently described molten globule (Kuwajima, 1989).

Results

Creation of a synthetic gene for *E. coli* RNase H

Using synthetic oligonucleotides, we have created a completely synthetic gene encoding the sequence of all 155 residues of *E. coli* RNase H with all 3 native cysteines replaced by alanines.

These amino acid substitutions were made in order to avoid complicating disulfide chemistry and to simplify later biochemical manipulations with the resulting protein. Previous reports indicated that such a variant protein retains the native level of activity (Kanaya et al., 1990) and hence, we assumed, is folded; we have confirmed this. For convenience, we refer to this variant sequence as simply RNase H. The modular nature of this gene construction provides conveniently placed restriction sites for cassette mutagenesis. Details on the construction of the plasmid and gene sequence are available upon request. Purification of the overexpressed gene product and confirmation of the final protein are described in detail in the Materials and methods.

Characterization of the native protein

The folding/unfolding transition of our recombinant RNase H was monitored by both CD and tryptophan fluorescence. As expected from previous structural data on the protein (X-ray and NMR), the far-UV CD spectrum at pH 8.0 is consistent with that of a mixed $\alpha + \beta$ protein (Fig. 2). The large negative signal near 222 nm, which reflects secondary structure ($-\theta_{222}$), allows us to follow folding by monitoring this single wavelength.

Thermal denaturation monitored in a buffered aqueous solution was found to be irreversible. However, when carried out in the presence of 0.5 M GuHCl, the process was >95% reversible. Note that the chemical denaturation studies described below demonstrate that the protein is completely folded under these mildly denaturing conditions. The T_m for this temperature-induced unfolding process, assuming a 2-state transition (see Materials and methods), is $55 \pm 0.2^\circ\text{C}$ (Fig. 3). Reversibility was also obtained when carried out at pH 3.0 in the absence of any added GuHCl; analysis of these data yielded similar results (data not shown).

Chemical denaturation using both urea and GuHCl was found to be completely reversible. Urea denaturation profiles were monitored by both CD and tryptophan fluorescence. These data were subjected to a 6-parameter, nonlinear least-squares anal-

Fig. 2. The far-UV CD spectrum of the native form of RNase H (●) (2 mM potassium phosphate, 5 mM potassium fluoride, pH 8.0, 25°C , $8.17 \mu\text{M}$ protein, pathlength 1 mm), and the partly folded acid form of RNase H (▲) (10 mM phosphoric acid, 5 mM potassium fluoride, pH 1.0, 25°C , $10.9 \mu\text{M}$ protein).

Fig. 3. Thermal denaturation of the native form of RNase H monitored by CD in the presence of 0.5 M GuHCl. The closed circles represent data obtained during heating, and the crosses represent data obtained during cooling. Raw data were converted to the fraction of folded protein as described in the Materials and methods (20 mM potassium phosphate, pH 8.0, 50 mM potassium chloride, 15 μ g/mL [0.86 μ M] protein).

ysis in order to determine the extrapolated free energy of folding in the absence of denaturant, $\Delta G(\text{H}_2\text{O})$, and the midpoint of the transition curve, C_m (see Materials and methods). Using the results of this analysis, the data were converted into fraction folded as shown in Figure 4. The 2 probes, CD and tryptophan fluorescence, yielded coincident transition curves at pH 8.0, suggesting that the process is indeed 2-state. The tryptophan fluorescence and CD studies using urea resulted in a $\Delta G(\text{H}_2\text{O}) = -7.6 \pm 1.3$ kcal/mol, $C_m = 3.93$ M, and $\Delta G(\text{H}_2\text{O}) = -7.4 \pm 1.0$ kcal/mol, $C_m = 3.84$ M, respectively (Fig. 4A) (reported errors were derived from multiple measurements). Denaturation by GuHCl monitored by CD (Fig. 4B) resulted in a $\Delta G(\text{H}_2\text{O}) = -7.2 \pm 0.7$ kcal/mol and a $C_m = 1.60$ M. These extrapolated values obtained for different probes and different denaturants are in very good agreement (Table 1). This close agreement helps validate the 2-state assumption, and therefore these data show no evidence for an equilibrium folding intermediate at pH 8.0.

Characterization of an acid equilibrium intermediate state

Protein stability as a function of pH was investigated by monitoring the CD signal at 222 nm (Fig. 5). This pH titration curve reveals an equilibrium intermediate state present below $\sim$ pH 1.8. Similar results were seen by fluorescence (data not shown). At around neutral pH, the native protein displays a CD signal at 222 nm of $\sim -10,000$ deg cm^2/dmol , whereas at low pH, the signal appears to plateau to a new value of $\sim -7,000$ deg cm^2/dmol . This value for $[\theta]_{222}$ is much larger than that expected for a completely denatured protein that has a signal at 222 nm close to 0 deg cm^2/dmol . Figure 2 shows the far-UV CD spectrum of this acid state. This spectrum is inconsistent with expectations from a completely denatured protein and suggests the presence of secondary structure at this low pH. This spectrum is also distinctly different from the spectrum of the native state.

The low pH form of the protein can be further denatured by either increased temperature or the addition of chemical denaturant. Figure 6A shows this thermal denaturation as monitored

Fig. 4. A: Urea denaturation of RNase H as monitored both by the CD signal at 222 nm ($\blacktriangle$) and by tryptophan fluorescence ($\circ$) (20 mM potassium phosphate, 50 mM potassium chloride, pH 8.0, 25 $^\circ\text{C}$, 40 μ g/mL protein). B: GuHCl denaturation of RNase H as monitored by the CD signal at 222 nm (20 mM potassium phosphate, 50 mM potassium chloride, pH 8.0, 25 $^\circ\text{C}$, 30 μ g/mL protein). Raw data were converted to fraction folded as described in the Materials and methods.

by following the CD signal at 222 nm. The transition is very broad. Although broad, there is some minimal apparent cooperativity in this transition (see Discussion). Somewhat surprisingly, this transition is completely reversible.

Chemical denaturation of the acid state monitored by both CD and tryptophan fluorescence is shown in Figure 6B. The transition is quite broad and the general shape is very similar when monitored by the 2 structural probes. These results suggest that the acid state contains residual structure both at the level of secondary structure (CD) and tertiary structure (as measured by the change in hydrophobic environment of the tryptophans). The protein contains 6 tryptophan residues, most of which reside near the so-called "basic-protrusion" area (see Discussion). Near-UV CD was used to further elucidate the level of tertiary packing within the core of the acid state. Figure 7

Table 1. Free energy of unfolding of ribonuclease H

	$\Delta G(\text{H}_2\text{O})$ (kcal/mol)	C_m (M)
CD(GuHCl)	-7.2 ± 0.7	1.60
CD(Urea)	-7.4 ± 1.0	3.84
Trp fluorescence(Urea)	-7.6 ± 1.3	3.93

Fig. 5. pH dependence of the CD signal at 222 nm for RNase H (10 mM sodium citrate, 10 mM sodium borate, 10 mM sodium phosphate, 50 mM potassium chloride, 30 μ g/mL [1.7 μ M] protein, 25 $^{\circ}$ C).

shows that this aromatic CD spectrum of the acid state more closely resembles that of the unfolded state than the well-packed core of the native state.

The acid intermediate is independent of protein concentration from 15 to 60 μ g/mL RNase H as monitored by CD (Fig. 8). These data indicate that the intermediate has about 70% of the signal of the native protein. In addition, they suggest that the structure-forming species is monomeric and not the result of intermolecular interactions or aggregation.

Equilibrium and velocity sedimentation experiments also indicate that RNase H remains monomeric both in the native state and acid state at a concentration of 240 μ g/mL. Equilibrium sedimentation experiments indicate a single species with a molecular weight of $\sim$ 16.9 kDa (pH 8.0) and 17.6 kDa (pH 1.0). The molecular weight of the protein as determined by mass spectrometry is 17.5 kDa.

Velocity sedimentation experiments carried out at 60,000 rpm indicate that the sedimentation coefficient of the native state is 1.79S, whereas that of the acid state is 1.65S. Again, this is strong evidence that the acid state is monomeric: only a single species is seen sedimenting for both the native and acid state. In addition, the acid state sediments with a lower value than that of the native state, indicating that aggregation is not taking place in the acid state.

The velocity sedimentation data also demonstrate that the acid state is not as compact as the native state: it has a slightly lower sedimentation coefficient. The sedimentation coefficient of the denatured state, corrected for density and viscosity differences, is significantly lower, with a value of 0.88S. Therefore, although a conformational change has occurred under acidic conditions, the change is not as drastic as that of the completely denatured state.

1-Anilinonaphthalene, 8-sulfonate (ANS) is a fluorescent dye that binds hydrophobic regions of proteins. Use of this fluorescent probe has been particularly helpful in the identification of equilibrium intermediates termed molten globules. The probe binds and fluoresces intensely to the molten globule state presumably because the loose tertiary interactions present in molten globules enable the probe to find exposed hydrophobic regions. Binding and fluorescence of ANS in unfolded proteins has a much lower quantum yield due to quenching by the surrounding solvent. ANS fluorescence spectra of the native, urea-denatured, and acid state of RNase H are shown in Figure 9.

Fig. 6. **A:** Thermal denaturation of the acid state of RNase H as monitored by following the CD signal at 222 nm (10 mM phosphoric acid, 5 mM potassium fluoride, pH 1.0, 30 μ g/mL [1.7 μ M] protein). **B:** Urea denaturation of the acid state of RNase H monitored by tryptophan fluorescence (upper) and the CD signal at 222 nm (lower) (20 mM phosphoric acid, 50 mM potassium chloride, pH 1.0, 40 μ g/mL [2.3 μ M] protein, 25 $^{\circ}$ C).

They indicate that the acid state of the protein binds ANS and fluoresces to a much greater extent than either the native or denatured protein, in accord with recent experiments on other molten globules.

Discussion

We have demonstrated that a variant of *E. coli* RNase H, in which all 3 free cysteines have been replaced by alanine, is a good model for studies on protein folding. This small single-domain protein has substantial secondary structure as measured by CD and contains 6 tryptophans that allow folding to be monitored by fluorescence. The protein folds reversibly and is quite stable under standard aqueous conditions. Thermal denaturation studies indicate a thermal midpoint of 55 $^{\circ}$ C, whereas chemical denaturation studies result in an extrapolated free energy of sta-

Fig. 7. Near-UV CD spectra for the native (lower tracing, closed circles) (20 mM potassium phosphate, 50 mM potassium chloride, pH 8.0, 25 °C), acid (center tracing, solid line) (20 mM potassium phosphate, 85 mM potassium chloride, pH 1.0, 4 °C), and unfolded (upper tracing, open circles) states (20 mM potassium phosphate, 50 mM potassium chloride, pH 8.0, 3.0 M GuHCl, 25 °C). Data were taken in a 1-cm-pathlength cuvette.

bilization in water, $\Delta G(\text{H}_2\text{O})$, of 7.2 kcal/mol (25 °C). Equilibrium denaturation studies at neutral pH were consistent with a 2-state transition and showed no evidence for an equilibrium folding intermediate.

pH-induced unfolding studies uncovered a stable equilibrium intermediate at low pH. The far-UV CD spectrum for this acid intermediate is distinctly different from that of the native state. In particular, the acid state appears to contain a higher percentage of α -helix as seen by the double minima at 222 nm and 205 nm (Fig. 2). The change from the native state to the acid state occurs in a sharp cooperative transition with a midpoint around pH 1.8. This pH transition is completely reversible at the concentrations studied. Further evidence for secondary structure at this low pH is demonstrated by the fact that the CD signal or secondary structure can be eliminated by addition of heat. Therefore, the acid state is not simply the unfolded form of the protein. Although very broad, the thermal unfolding curve ap-

Fig. 8. The concentration dependence of the CD signal at 222 nm for both the native (closed circles) and acid states (open circles) of RNase H at 25 °C (native state: 2 mM potassium phosphate, 5 mM potassium chloride, pH 8.0; acid state: 5 mM phosphoric acid, 5 mM potassium chloride, pH 1.0).

Fig. 9. ANS fluorescence spectra of the acid (solid line), urea-denatured (dashed line), and native states (broken line) of RNase H. Excitation wavelength = 405 nm; 1.71 μM protein, 0.427 mM ANS (acid state: 5 mM phosphoric acid, 5 mM potassium chloride, pH 1.0; denatured state: 6 M urea, 5 mM phosphoric acid, 5 mM potassium chloride, pH 1.0; native state: 2 mM potassium phosphate, 5 mM potassium chloride, pH 8).

pears to be somewhat cooperative, suggesting the possibility of a true transition between the acid state and the unfolded state of the protein. Such a broad thermal unfolding curve would be consistent with a well-folded protein with a T_m of ~ 25 °C. It will be particularly interesting to determine if one can observe any measurable calorimetric enthalpy with this transition.

Because we were able to follow chemical denaturation of the acid state by monitoring tryptophan fluorescence, at least some of the 6 tryptophans in RNase H must have different hydrophobic environments in the acid and denatured states. These data imply that the acid state must contain some tertiary interactions that bury its aromatic amino acids. The lack of a pronounced near-UV CD spectrum suggests that, although shielded from water in a hydrophobic environment, the aromatic amino acids are not rigidly fixed in tight tertiary interactions. These data, together with the far-UV CD data, imply that the acid state must contain both residual secondary structure (CD) and at least some loose tertiary interactions (fluorescence). Because all 6 tryptophans are located near the so-called basic protrusion area of the protein, this area may be playing an important role in the stability and/or structure of the acid state. Interestingly, the RNase H domain from human immunodeficiency virus (HIV) does not have this protrusion and also does not form an acid intermediate state (J. Keck & S. Marqusee, pers. comm.). The role of this basic protrusion is being explored further by site-specific mutagenesis.

ANS fluorescence experiments demonstrate that the acid state contains some accessible hydrophobic regions within a loosely collapsed structure. Clearly, the native and denatured states do not bind and fluoresce to the same extent as the acid state. Thus, although ANS cannot penetrate the tightly packed hydrophobic core of the native state, the acid intermediate, which is less compact than the native state, allows ANS to penetrate the protein and enables the dye to bind to hydrophobic regions of the protein. In addition, the residual structure present in the acid state prevents energy from being lost to solvent interactions that presumably occur in the denatured state.

Both the CD and centrifugation experiments suggest that the acid state is monomeric and not the result of intermolecular interactions or aggregation. The CD signal for both the acid state

(pH 1.0) and the native state (pH 8.0) are linear over the protein concentration range studied (0–3.4 μ M). Equilibrium and velocity sedimentation experiments ($\sim$ 0.4 mM) also indicate that RNase H is monomeric under acidic and native conditions. In addition, we have observed that the acid state does not aggregate easily at high concentrations (mM). This is quite different from other equilibrium intermediates, or molten globules, which usually aggregate—probably because the exposed hydrophobic regions of partly folded proteins interact, causing the proteins to fall out of solution. Because of this limitation, most structural information on molten globules have been obtained indirectly by using hydrogen–deuterium exchange under molten globule conditions, followed by quenching and ultimate NMR observation under native conditions (Baum et al., 1989; Hughson et al., 1990). Perhaps the extremely low pH of our acid state prohibits these intermolecular interactions from occurring. The apparently high solubility of our acid intermediate should make it amenable to direct structural studies by NMR.

The acid state discovered for RNase H is similar to other equilibrium intermediates that have been found under extreme solvent conditions for a variety of proteins such as α -lactalbumin (Kuwaitjima et al., 1976; Alexandrescu et al., 1993), cytochrome *c* (Jeng et al., 1990), and apomyoglobin (Griko et al., 1988; Hughson et al., 1990). These equilibrium intermediates have been termed molten globules. The molten globule state has been operationally defined by the following characteristics (Kuwaitjima, 1989): (1) retention of some secondary structure; (2) tertiary interactions are absent or weak at best; and (3) some level of compactness is retained. Characterization of the acid state of RNase H demonstrates that it has all of these properties of a molten globule as measured by CD, sedimentation, and ANS binding.

There is mounting evidence that transient intermediates similar to molten globules are populated early in the kinetic pathway of protein folding. Kinetic data derived from stopped-flow studies reveal transient species that bind ANS very early in folding, even for proteins that have no known equilibrium intermediate (Ptitsyn et al., 1990). Far-UV CD studies of α -lactalbumin also suggest that the acid state of a protein has properties similar to transient intermediates in the kinetic pathway (Ikeguchi et al., 1986). In addition, high-resolution NMR data on the protein apomyoglobin indicate that residues with structure in the molten globule state correspond to areas in the protein that are protected early in the folding process from hydrogen–deuterium exchange as measured by NMR (Jennings & Wright, 1993). Given the solubility, reversibility, and relative ease in manipulating the acid intermediate of RNase H, it will be particularly important to determine if a state similar to the acid state of RNase H is transiently populated on the kinetic folding pathway.

Materials and methods

Construction of a synthetic gene for RNase H

Plasmid pSM101 is a derivative of the T7 expression plasmid pAED4 (Doering, 1992). This plasmid carries a synthetic gene for *E. coli* RNase H with all 3 cysteines replaced by alanine. Expression of the gene is under the control of the T7 promoter. The gene sequence was engineered to distribute unique restriction sites throughout the sequence. The plasmid was constructed from the following fragments: (1) the ClaI–NdeI fragment of pAED4, which bears the T7 promoter; (2) 13 synthetic oligo-

nucleotides that overlap and anneal to form the coding sequence of the RNase H variant.

Protein purification

The vector pSM101 was transformed into the T7 expression strain BL21 (DE3), plys S (Novagen). Cells were grown in standard Luria broth with 100–200 μ g/mL of ampicillin at 37 °C. Overexpression of the protein was induced by addition of isopropyl- β -thiogalactoside (1 mM) to log-phase cultures. Three hours after induction, cells were harvested and sometimes stored as pellets at -70 °C until needed.

Cell pellets were thawed and resuspended in ice-cold buffer A (50 mM Tris, pH 8.0, 20 mM NaCl, 10 mM MgCl₂, 0.5 mM EDTA, 1 mM β -mercaptoethanol). Cells were lysed by mild sonication and cell debris was removed by centrifugation. The protein was determined to reside in the soluble fraction by SDS-PAGE and Coomassie blue staining.

Purification was achieved in 2 chromatographic steps. After cell lysis, the soluble fraction was loaded directly onto a heparin-agarose column (Pharmacia). Previous experiments have shown that heparin specifically inhibits *E. coli* RNase H activity (Dirksen & Crouch, 1981). The protein was eluted by the addition of 5% buffer B (buffer A + 2 M KCl). Fractions containing RNase H were pooled and concentrated (centriprep 10, Amicon) to a volume of $\sim$ 1.5 mL. The concentrated sample was then size fractionated over a 120-mL HiLoad 16/60 Sephacryl S-100 fast protein liquid chromatography column (Pharmacia) (20 mM potassium phosphate, 50 mM KCl) at 0.5–1.0 mL/min. Purified protein was judged to be $>95\%$ homogeneous following SDS-PAGE and staining with Coomassie blue. This purification typically yielded 10–35 mg of pure protein per liter of cells. Protein was either stored frozen at -20 °C until needed or dialyzed extensively against 50 mM ammonium bicarbonate, lyophilized, and resuspended into the designated buffer before use. No difference in CD signal was observed over months of storage.

Protein concentration was determined by UV absorption spectroscopy (OD_{280} at 1 mg/mL = 2.02 [Kanaya et al., 1990]). The molecular weight of the protein was verified by mass spectroscopy to be 17,502 Da, indicating production of the full-length protein retaining the initiator methionine.

CD measurements

CD measurements were made on an Aviv 62DS spectropolarimeter equipped with a peltier temperature-controlled sample holder and either a 10-mm- or 1-mm-pathlength cuvette.

Protein unfolding was assayed by monitoring the ellipticity at 222 nm as a function of pH, temperature, and urea or GuHCl concentration. Chemical denaturation experiments were performed at 25 °C using protein samples at a concentration of 30–40 μ g/mL in a buffer consisting of 20 mM potassium phosphate, 50 mM potassium chloride (pH 8.0), or 20 mM phosphoric acid, 50 mM potassium chloride (pH 1.0). For thermal denaturation, samples were equilibrated for at least 2 min at each temperature, and the signal was averaged over a 1-min interval. In order to achieve reversibility, thermal denaturations on the native state were carried out either at pH 8.0 in the presence of 0.5 M GuHCl, or in the absence of GuHCl at pH 3.3 (10 mM sodium citrate, 10 mM sodium borate, 10 mM sodium phosphate). Con-

ditions for thermal denaturation on the acid state were 10 mM phosphoric acid, 5 mM potassium fluoride, pH 1.0.

Tryptophan fluorescence measurements

Urea denaturation curves were monitored by tryptophan fluorescence using a Perkin Elmer model LS-50B luminescence spectrometer with a pathlength of 1 cm at ambient temperature. Excitation was at 295 nm and emission was monitored at 340 nm.

Analysis of chemical denaturation and free energy determination

The free energy of unfolding was determined using a 2-state model ($N \rightleftharpoons D$) and a linear dependence of free energy on denaturant concentration (Santoro & Bolen, 1988):

$$\begin{aligned}\Delta G_u &= \Delta G(\text{H}_2\text{O}) - m * [\text{urea}] \\ &= -RT \ln K_u = -RT \ln \left(\frac{y_N - y}{y - y_D} \right),\end{aligned}$$

where $\Delta G(\text{H}_2\text{O})$ is the extrapolated free energy in the absence of denaturant, y is the observed CD signal, and y_N and y_D are the ellipticities characteristic of the native and denatured states. The constant m measures the dependence of ΔG_u on denaturant concentration. Multiple data sets were collected and fit using the program NONLIN (Santoro & Bolen, 1988) to the following equation:

$$y = \frac{(y_f + m_f * [\text{denaturant}]) + (y_u + m_u * [\text{denaturant}]) * (\exp[-\Delta G_u/RT])}{1 + \exp(-\Delta G_u/RT)}.$$

In this equation, ΔG_u is a function of $\Delta G(\text{H}_2\text{O})$, m , and urea as described above, and y_f , m_f and y_u , m_u are the 0-denaturant intercepts and slopes for the native and denatured baselines, respectively. Reported uncertainties are experimental errors derived from multiple experiments.

Analysis of thermal denaturation and T_m

The midpoint temperature, or T_m , for thermal denaturation was analyzed in a similar fashion to the chemical denaturation profiles using a 2-state assumption and the following model:

$$\begin{aligned}\Delta G &= \left(\frac{T_m - T}{T_m} \Delta H \right) - (T_m - T) \Delta C_p + T \Delta C_p \ln \frac{T_m}{T} \\ &= -RT \ln K_u = -RT \ln \left(\frac{y_N - y}{y - y_D} \right).\end{aligned}$$

Data were fit using the program NONLIN to the following equation:

$$y = \frac{(y_f + m * T) + (y_u + m_u * T) * (\exp[-\Delta G_u/RT])}{1 + \exp(-\Delta G_u/RT)}.$$

Reported uncertainties are experimental errors derived from multiple experiments.

Analytical ultracentrifugation

Sedimentation equilibrium and sedimentation velocity experiments were carried out on a Beckman Optima XL-A ultracentrifuge, using 20,000 and 60,000 rpm, respectively, and a radius (r) of 5.80–7.20 cm. All runs were carried out at 15 °C with a protein concentration of 240 $\mu\text{g/mL}$. Conditions for analysis of the native state were 20 mM potassium phosphate, 100 mM NaCl, pH 8.0; those for the acid state were 20 mM phosphoric acid, 100 mM NaCl, pH 1.0. Denatured state conditions were 20 mM potassium phosphate, 50 mM potassium chloride, 3 M GuHCl, pH 8.0.

Data analyses were performed with the standard software provided by Beckman. Sedimentation equilibrium data were evaluated assuming a 2-component system (solvent and solute), fitting a model curve to the observed distribution. Estimated values needed to carry out these studies were a solvent density of 1.004 g/cm^3 and a partial specific volume 0.717 cm^3/g (note that equilibrium studies were only performed on the native and acid states). Measurements of sedimentation velocity utilized the movement of the inflection point of the boundary of protein as a function of time (correlation coefficients were greater than 0.994).

ANS fluorescence measurements

ANS was used as a qualitative probe for the molten globule state. Experiments were carried out with a Hitachi F-4500 fluorimeter using a pathlength of 1 cm at ambient temperature. Excitation wavelength was 405 nm and emission was monitored from 425 to 675 nm. All spectra were blank subtracted. The molar ratio of ANS:RNaseH was 250:1 (Semisotnov et al., 1991).

Acknowledgments

This work was supported in part by a grant from the Lucille P. Markey Charitable Trust. We thank Y.-R. Yang and H. Schachman for help with the analytical ultracentrifuge; D. King for mass spectroscopy; A. Glazer and C.R. Matthews for advice and discussion; and A. Chamberlain, J. Keck, and S. Tsutakawa for helpful discussion and critical reading of the manuscript.

References

- Alexandrescu AT, Evans PA, Pitkeathly M, Baum J, Dobson CM. 1993. Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: A two-dimensional NMR study. *Biochemistry* 32: 1707–1718.
- Baum J, Dobson CM, Evans PA, Hanley C. 1989. Characterization of a partly folded protein by NMR methods: Studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry* 28:7–13.
- Crouch RJ, Dirksen M. 1982. Ribonucleases H. In: Linn S, ed. *Nucleases*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Davies J, Hostomska Z, Hostomsky Z, Jordan SR, Matthews DA. 1991. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* 252:88–95.
- Dirksen ML, Crouch RJ. 1981. Selective inhibition of RNase H by dextran. *J Biol Chem* 256:11569–11573.
- Doering DS. 1992. *Functional and structural studies of a small f-actin binding domain*. Cambridge, Massachusetts: MIT Press.
- Griko YV, Privalov PL, Venyaminov SY, Kutysheko VP. 1988. Thermodynamic study of the apomyoglobin structure. *J Mol Biol* 202:127–138.
- Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249:1544–1548.
- Ikeguchi M, Kuwajima K, Mitani M, Sugai S. 1986. Evidence for identity between the equilibrium unfolding intermediate and a transient folding

- intermediate: A comparative study of the folding reactions of alpha-lactalbumin and lysozyme. *Biochemistry* 25:6965-6972.
- Ishikawa K, Nakamura H, Morikawa K, Kanaya S. 1993. Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry* 32:6171-6178.
- Jeng MF, Englander SW, Elove GA, Wand AJ, Roder H. 1990. Structural description of acid-denatured cytochrome *c* by hydrogen exchange and 2D NMR [published erratum appears in *Biochemistry* 30:2988, Mar 19, 1991]. *Biochemistry* 29:10433-10437.
- Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin [see comments]. *Science* 262:892-896.
- Johnson MS, McClure MA, Feng DF, Gray J, Doolittle RF. 1986. Computer analysis of retroviral pol genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc Natl Acad Sci USA* 83:7648-7652.
- Kanaya S, Katsuda C, Kimura S, Nakai T, Kitakuni E, Nakamura H, Katayanagi K, Morikawa K, Ikehara M. 1991. Stabilization of *Escherichia coli* ribonuclease H by introduction of an artificial disulfide bond. *J Biol Chem* 266:6038-6044.
- Kanaya S, Kimura S, Katsuda C, Ikehara M. 1990. Role of cysteine residues in ribonuclease H from *Escherichia coli*. Site-directed mutagenesis and chemical modification. *Biochem J* 271:59-66.
- Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Nakamura H, Ikehara M, Matsuzaki T, Morikawa K. 1992. Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J Mol Biol* 223:1029-1052.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946-950.
- Kuwajima K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins Struct Funct Genet* 6:87-103.
- Kuwajima K, Nitta K, Yoneyama M, Sugai S. 1976. Three-state denaturation of alpha-lactalbumin by guanidine hydrochloride. *J Mol Biol* 106:359-373.
- Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI. 1990. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett* 262:20-24.
- Santoro MM, Bolen DW. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27:8063-8068.
- Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. 1991. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* 31:119-128.
- Yamazaki T, Yoshida M, Kanaya S, Nakamura H, Nagayama K. 1991. Assignments of backbone ¹H, ¹³C, and ¹⁵N resonances and secondary structure of ribonuclease H from *Escherichia coli* by heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* 30:6036-6047.
- Yang W, Hendrickson W, Crouch RJ, Satow Y. 1990. Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* 249:1398-1405.