BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 5167–5175 5167
DOI: 10.1021/bi1001097

Identification of Residual Structure in the Unfolded State of Ribonuclease H1 from the Moderately Thermophilic *Chlorobium tepidum*: Comparison with Thermophilic and Mesophilic Homologues[†]

Kathleen Ratcliff[§,∥] and Susan Marquesee*[,‡,§,∥]

[‡]*Department of Molecular and Cell Biology,* [§]*Biophysics Graduate Group, and* [∥]*Institute of Quantitative Biosciences (QB3)—Berkeley, University of California, Berkeley, Berkeley, California 94720*

Received January 24, 2010; Revised Manuscript Received May 17, 2010

ABSTRACT: Ribonucleases H from organisms that grow at different temperatures demonstrate a variable change in heat capacity upon unfolding ($\Delta C°_P$) [Ratcliff, K., et al. (2009) *Biochemistry 48*, 5890−5898]. This $\Delta C°_P$ has been shown to correlate with a tolerance to higher temperatures and residual structure in the unfolded state of the thermophilic proteins. In the RNase H from *Thermus thermophilus*, the low $\Delta C°_P$ has been shown to arise from the same region as the folding core of the protein, and mutagenic studies have shown that loss of a hydrophobic residue in this region can disrupt this residual unfolded state structure and result in a return to a more mesophile-like $\Delta C°_P$ [Robic, S., et al. (2002) *Protein Sci. 11*, 381−389; Robic, S., et al. (2003) *Proc. Natl. Acad. Sci. U.S.A. 100*, 11345−11349]. To understand further how residual structure in the unfolded state is encoded in the sequences of these thermophilic proteins, we subjected the RNase H from *Chlorobium tepidum* to similar studies. Analysis of new chimeric proteins reveals that like *T. thermophilus* RNase H, the folding core of *C. tepidum* RNase H plays an important role in the unfolded state of this protein. Mutagenesis studies, based on both a computational investigation of the hydrophobic networks in the core region and comparisons with similar studies on *T. thermophilus* RNase H, identify new residues involved in this residual structure and suggest that the residual structure in the unfolded state of *C. tepidum* RNase H is more restricted than that of *T. thermophilus*. We conclude that while the folding core region determines the thermophilic-like behavior of this family of proteins, the residue-specific details vary.

The mechanisms by which proteins from thermophilic organisms are able to function at extreme temperatures are not well understood. Identifying a more complete set of the rules governing this thermophilicity would further our understanding of proteins from thermophilic organisms and would contribute to our general knowledge about protein stability and energetics. Studies on sets of homologous proteins from thermophilic and mesophilic organisms show that proteins with the same function have nearly identical folds. Furthermore, the sequence and structural similarities between such homologous proteins are often indistinguishable from those between a pair of homologous proteins from mesophiles (*1*−*6*). The ribonucleases H from *Escherichia coli*, *Chlorobium tepidum*, and *Thermus thermophilus* form a set of well-studied homologous proteins with very different thermal stability profiles. They are small, single domain proteins with $\alpha + \beta$ architecture; the sequence identity between all three is around 50% (*7*). All three fold into a virtually indistinguishable RNase H[1]-like fold (*1*, *8*). These RNases H provide a good system for examining the subtle changes in sequence that lead to large changes in thermodynamic behavior.

The energetics of these RNases H from *T. thermophilus*, *C. tepidum*, and *E. coli* (TthRNH*, CtepRNH*, and EcoRNH*, respectively, where the asterisk indicates a cysteine-free variant) have been studied extensively via thermal and chemical denaturation (*9*). The basic folding and unfolding kinetics are also known for all three (*9*, *10*). In the case of TthRNH* and EcoRNH*, the energy landscapes have also been evaluated by hydrogen exchange (*11*−*13*). The stability curves ($\Delta G°_{unf}$ vs temperature profiles) show that, in addition to a higher melting temperature (T_m), TthRNH* has a dramatically lower change in heat capacity upon unfolding ($\Delta C°_P$) and a higher overall stability ($\Delta G°_{unf}$) than EcoRNH* (*11*). Recently, we carried out a similar thermodynamic analysis on the bacterial RNase H from *C. tepidum*. *C. tepidum*, a moderate thermophile, has a growth temperature (48 °C) between those of *E. coli* (37 °C) and *T. thermophilus* (68 °C). Protein stability curves revealed that CtepRNH* has approximately the same stability at 25 °C and melting temperature as EcoRNH* while maintaining the unusually low $\Delta C°_P$ observed for the thermophilic TthRNH* (*9*).

Both site-directed mutagenesis and studies of chimeric proteins implicate the folding core of TthRNH* as the region responsible for its low $\Delta C°_P$. Using EcoRNH* and TthRNH*, two chimeric proteins were created, each composed of the folding core of one RNase H and the periphery of the other: ECTO (*E. coli* core,

[†]Research in the laboratory of S.M. is supported by Grant GM50945 from the National Institutes of Health.

*To whom correspondence should be addressed. Tel: (510) 642-7678. Fax: (510) 666-2765. E-mail: marqusee@berkeley.edu.

[1]Abbreviations: CD, circular dichroism; RNase H, ribonuclease H; $\Delta C°_P$, change in heat capacity upon unfolding; $\Delta G°_{unf}$, change in free energy upon unfolding ($G_u - G_f$); CtepRNH, ribonuclease H from *C. tepidum*; EcoRNH, ribonuclease H from *E. coli*; TthRNH, ribonuclease H from *T. thermophilus*; CtepRNH*, cysteine-free version of the ribonuclease H from *C. tepidum*; EcoRNH*, cysteine-free version of the ribonuclease H from *E. coli*; TthRNH*, cysteine-free version of the ribonuclease H from *T. thermophilus*; CCEO, chimeric ribonuclease H with *C. tepidum* core and *E. coli* outside; CCTO, chimeric ribonuclease H with *C. tepidum* core and *T. thermophilus* outside.

Published on Web 05/21/2010

T. thermophilus outside) and TCEO (*T. thermophilus* core, *E. coli* outside). The boundaries between these regions (core/periphery) were defined through hydrogen exchange studies and an algorithm called RAFT (*10*, *12*, *14*). RAFT is a structure-based algorithm that predicts autonomous folding units by identifying contiguous regions of a protein with a high density of intra-region contacts and a minimum of inter-region contacts. The core regions of EcoRNH* and TthRNH* were identified as high scoring regions based on the RAFT algorithm. Both ECTO and TCEO formed well-folded and active RNases H; the X-ray structure of TCEO showed an RNase H fold indistinguishable from the other RNases H. Thermodynamic analyses of these two chimeras revealed that TCEO is destabilized as compared to TthRNH* yet maintains the low ΔC°_P associated with the thermophile, while ECTO is also destabilized but maintains the high ΔC°_P of the mesophile (*15*). A single point mutation (I53D) in the core region can alter the ΔC°_P of TthRNH* dramatically, bringing it close to the predicted higher ΔC°_P that is present in the mesophilic homologue, while the same mutation does not alter the ΔC°_P of EcoRNH* (*16*). Furthermore, the core regions of both EcoRNH* and TthRNH* have been shown to fold autonomously (*17*). These observations all lead to the hypothesis that the folding core is crucial to the thermodynamic characteristics of these proteins and, importantly, the ability to fold and function at elevated temperatures.

The major contributor to the ΔC°_P ($C^\circ_{Pu} - C^\circ_{Pf}$) associated with protein unfolding is thought to be the change in solvent exposure of nonpolar residues (*18*). The above mutagenic studies, together with differential calorimetric studies, suggest that the low ΔC°_P in TthRNH* arises from the unfolded state (C°_{Pu}) (*16*, *19*); this residual structure in the unfolded state resides in the core region of the protein, and the mutation I53D appears to disrupt this residual structure.

To explore this hypothesis further, we wondered if the core region of CtepRNH* also encodes the unusually low ΔC°_P observed for this protein. Like the other two proteins, the folding kinetics of CtepRNH* reveal an early folding intermediate, which, by analogy, is likely composed of the same core region of the protein (*9*). To test this hypothesis, we have created chimeric proteins similar to those previously studied: CCEO (*C. tepidum* core, *E. coli* outside) and CCTO (*C. tepidum* core, *T. thermophilus* outside). Indeed, we find that the core region of CtepRNH* appears to be responsible for the low ΔC°_P. However, unlike in TthRNH*, we find that mutations at residue 53 do not affect the ΔC°_P of CtepRNH*. Instead, mutations at a different site (residue 56) implicate a nearby region as a crucial component of this residual structure. Furthermore, mutations in TthRNH* at this alternative site have an even greater impact on the ΔC°_P than do mutations at residue 53. Remarkably, site-specific mutagenesis experiments designed based on the native structure have allowed us to interrogate the nature of the unfolded state of a protein under native conditions. Our results suggest that modulation of the core region of RNase H plays an important role in its adaptation to different thermal niches. While the same core region seems to be important in both proteins, the exact method by which each enzyme encodes this residual structure in the unfolded state differs. There does not appear to be a unique amino acid sequence that defines this unfolded state property.

EXPERIMENTAL PROCEDURES

Construction of CCEO and CCTO. The core region of the ribonuclease H gene from *C. tepidum* (residues 43−121) was amplified by PCR using primers designed to add overlapping sequences to either *E. coli* or *T. thermophilus* RNase H* (*15*, *20*). The periphery regions of *E. coli* and *T. thermophilus* RNase H* were amplified in two parts from plasmid DNA (pSM101 and pJH109, respectively (*11*)) using primers designed to add overlapping sequences to either an NDE-1 restriction site and the *C. tepidum* core region or the *C. tepidum* core region and a KPN-1 restriction site (the beginning and ending fragments, respectively). The fragments were then annealed and amplified to make two complete chimeric RNase H genes, which were individually inserted into a pAED4 vector (*21*). Both strands of the coding region in the resulting plasmids (pKR401 and pKR402) were sequenced and found to be correct. Plasmids encoding CCEO and CCTO were transformed into *E. coli* BL21(DE3) pLysS cells (Novagen), and the cells were grown at 37 °C in Luria broth with 200 μg/mL ampicillin and 25 μg/mL chloramphenicol. Expression was induced at an OD_{600} of 0.6 by the addition of 1 mM IPTG, and cells were grown for an additional 3 h before harvesting by centrifugation.

Purification of CCEO and CCTO. Purification of CCEO and CCTO was carried out using a standard RNase H purification method. Briefly, cell pellets were resuspended in 50 mM Tris, pH 8.0, 20 mM NaCl, and 0.1 mM EDTA and lysed by sonication. The soluble fraction was loaded onto a HiTrap heparin HP column from GE Healthcare and eluted with a linear NaCl gradient (0.02−0.8 M NaCl). Fractions containing CCEO and CCTO were pooled and diluted 1:1 with water, and the pH of the resulting sample was adjusted to 5.5 with concentrated HCl. The protein sample was then loaded onto a Source 15S column (in 20 mM NaOAc, pH 5.5, 200 mM NaCl, and 0.1 mM EDTA) and eluted with a linear NaCl gradient (0.2−0.6 M NaCl). Fractions containing CCEO or CCTO were judged to be >98% pure by SDS−PAGE, pooled, and stored at −80 °C. The final masses of each protein were confirmed by mass spectrometry and found to be correct to within 2 amu.

Construction of C. tepidum RNase H Point Mutants*. The gene for *C. tepidum* RNase H was mutated at positions 53 and 56 using the QuikChange (Stratagene) site-directed mutagenesis protocol. The coding strands of all resulting mutant genes (I53A, I53D, L56A, L56D, and L56S) were sequenced and found to be correct. Each plasmid was transformed into *E. coli* BL21(DE3) pLysS cells (Novagen) in the presence of 200 μg/mL ampicillin and 25 μg/mL chloramphenicol, and the cells were grown at 37 °C in Luria broth. Expression of each protein was induced at an OD_{600} of 0.6 by the addition of 1 mM IPTG, and cells were grown for 3 h before harvesting by centrifugation.

Purification of C. tepidum RNase H Point Mutants*. All *C. tepidum* RNase H* I53 and L56 point mutants expressed insolubly in inclusion bodies, as was observed for all *E. coli* RNase H* I53 mutants. Each protein was purified as described previously (*22*).

Construction, Expression, and Purification of T. thermophilus RNase H L56S*. The gene for *T. thermophilus* RNase H* in an MBP fusion vector with a TEV protease-cleavable MBP+His tag was a gift from K. Connell. Residue L56 was mutated using the QuikChange site-directed mutagenesis protocol (Stratagene). The resulting *T. thermophilus* RNase H* L56S gene was sequenced and found to be correct. This plasmid was transformed into *E. coli* BL21(DE3) pLysS cells in the presence of 25 μg/mL chloramphenicol and 50 μg/mL kanamycin in Luria broth. Cells were induced for 3 h at 37 °C after reaching an OD_{600} of 0.6−0.8 by the addition of 1 mM IPTG. Cells were harvested

by centrifugation, and cell pellets were stored at −80 °C until purification.

T. thermophilus RNase H* L56S was purified using a three-column protocol. First, cells were lysed by sonication and spun down at 12000 rpm for 30 min. The supernatant was loaded onto a HisTrap nickel column (GE Healthcare) in 20 mM sodium phosphate, pH 8.0, 500 mM NaCl, 30 mM imidazole, 0.1 mM EDTA, and 1 mM TCEP. The MBP-TthRNH* L56S fusion protein was eluted in a step to buffer containing 500 mM imidazole. Protein was dialyzed overnight into low-imidazole buffer in the presence of 2 mg of TEV protease and then rerun through the same protocol on the HisTrap column. The flow-through was then dialyzed into 50 mM Tris, pH 8.0, 20 mM NaCl, and 0.1 mM EDTA. This sample was run over a HiTrap heparin column (GE Healthcare) and eluted with a linear gradient of NaCl. Fractions containing TthRNH* L56S were judged to be >98% pure by SDS−PAGE, pooled, and dialyzed into CD buffer for thermodynamic measurements. A sample of this and all other mutant proteins was determined to have the correct mass via mass spectrometry to within 1 amu.

GdmCl-Induced and Thermal Denaturation. Circular dichroism (CD) measurements were carried out on an Aviv 410 spectropolarimeter with a Peltier temperature-controlled cell holder. For all samples, each CD measurement is an average of the signal at 222 nm over 1 min in a 1 cm path length cuvette. Denaturation samples of all proteins were made to contain 50 μg/mL protein, 20 mM NaOAc (pH 5.5), 50 mM KCl, and the appropriate concentration of GdmCl. These samples were allowed to equilibrate between 1 and 3 days before the CD signal was measured. Twenty-five samples were used for each denaturation melt, and individual sets of samples were equilibrated at varying temperatures ranging from 0 to 55 °C. Free energies of unfolding ($\Delta G°_{unf}$) were determined from the GdmCl-induced denaturation of each protein at each temperature, assuming a two-state model and a linear relationship between $\Delta G°_{unf}$ and concentration of GdmCl (see below) (*23*). For thermal denaturation, the same buffer conditions lacking GdmCl were used. In these experiments, the CD signal was recorded every 3 °C, with a 5 min equilibration time and a 1 min averaging time at each temperature. At the end of the melt, reversibility was determined by returning to the initial temperature and comparison of the CD spectrum from 200 to 300 nm to the premelt spectrum. To determine the T_m for each protein, thermal melts were fit using a two-state model and the Gibbs−Helmholtz relationship between $\Delta G°_{unf}$ and temperature (*24*). The reported values represent the average of multiple experiments under identical conditions. The uncertainties reported for both $\Delta G°_{unf}$ and T_m are estimates based on the largest deviation from this mean. These uncertainties are slightly larger than the errors of the fit and therefore serve as the more conservative estimate.

Denaturation and Stability Curve Data Analysis. Due to the noted systematic and linear temperature-dependent variation in the *m*-value of each protein, the *m*-values were fit to a line, and the predicted *m*-value was used to refit each denaturant melt as described previously (*25*). The free energies of unfolding ($\Delta G°_{unf}$) for CCEO, CCTO, or any of the point mutants obtained from the denaturant melts as described above were plotted as a function of temperature, with a single point added for the T_m ($\Delta G°_{unf} = 0$) if the protein was sufficiently reversible to ensure accuracy of the T_m. These stability profiles ($\Delta G°_{unf}$ vs T) were then fit to a modified Gibbs−Helmholtz equation using the T_m as a reference temperature:

$$\Delta G°_{unf} = \Delta H° - T\left(\frac{\Delta H}{T_m}\right) + \Delta C°_P\left[T - T_m - T\ln\left(\frac{T}{T_m}\right)\right]$$

The data shown for each data point in a stability curve were the result of two to four repetitions of the same conditions, with corresponding error bars indicating the maximum observed difference from the mean. These uncertainties were then included during the fit of the stability curve to generate the $\Delta C°_P$ and an estimated error using a standard protocol in the Sigmaplot software analysis package (Systat Software Inc., San Jose, CA). These propagated errors are only slightly larger than the errors determined solely by the fit.

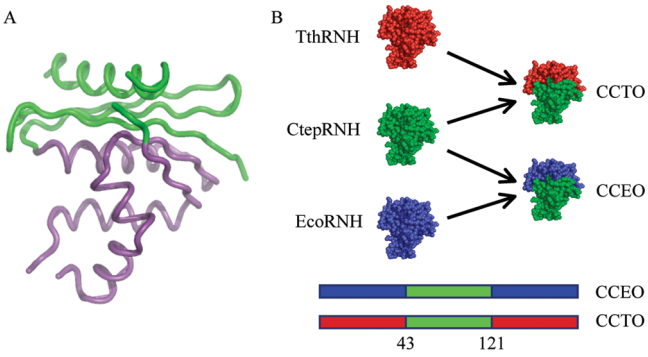


FIGURE 1: Structural representations of CtepRNH* and the designed chimeras. (A) Illustration of the boundaries between the core and periphery of CtepRNH*. The core region, highlighted in purple, is made up of residues 43−121, while the periphery, or outside, residues 1−42 and 122−145 are shown in green. (B) Schematic diagram of CCEO and CCTO. Residues taken from EcoRNH* are in blue, residues from TthRNH* are in red, and residues from CtepRNH* are in green.

RESULTS

Design and Initial Characterization of Chimeric RNases H. The core of CtepRNH* was identified based on homology modeling with both EcoRNH* and TthRNH*. The folding cores of EcoRNH* and TthRNH* were identified by hydrogen exchange studies coupled with a native structure contact-based algorithm (*14*, *15*). The CtepRNH* core region (residues 43−121) was found to have sequence identities of 62% and 54% and sequence similarities of 73% and 68% to TthRNH* and EcoRNH*, respectively. Two chimeric RNases H were designed to combine fragments of the moderately thermophilic CtepRNH* with the mesophilic EcoRNH* and the thermophilic TthRNH* (Figure 1). Both chimeras are composed of the core region of CtepRNH* with the remaining residues originating from either of the two other RNases H and are referred to as CCTO (*C. tepidum* core, *T. thermophilus* outside) and CCEO (*C. tepidum* core, *E. coli* outside). Modeling of the interfacial regions of each new protein suggests that no new charge-repulsion interactions have been introduced. In fact, the interfaces of each of these chimeras seem to avoid an unfavorable interaction created in the previously studied chimera ECTO (D66-E6) (*15*). Both chimeric proteins were overexpressed in *E. coli* and purified as previously described for TthRNH* (*11*).

Circular dichroism (CD) spectra indicate that both proteins are well folded and adopt secondary structures similar to those of their parent proteins (Figure 2A). RNase H activity was determined through a UV absorbance-based assay as previously

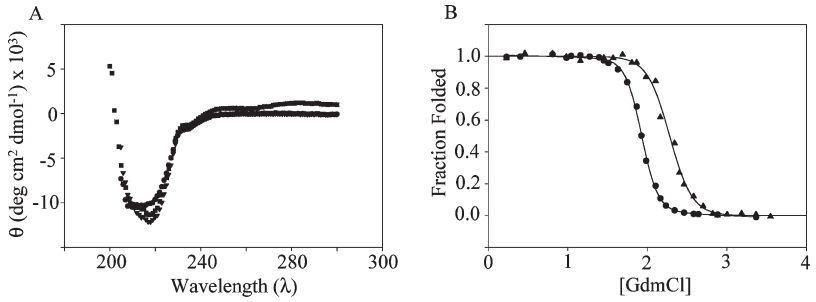


FIGURE 2: Biophysical properties of CCEO and CCTO determined using circular dichroism spectroscopy. (A) CD spectra of CtepRNH* (circles), CCEO (squares), and CCTO (triangles). Data were taken on 50 μg/mL protein at 25 °C, 50 mM KCl, 20 mM NaOAc, pH 5.5. (B) Denaturant melts of CCEO (circles) and CCTO (triangles) taken on 50 μg/mL protein at 25 °C, 50 mM KCl, 20 mM NaOAc, pH 5.5, and varying concentrations of GdmCl.

described (*15*, *26*), and both proteins showed similar activity to the parent protein that donated the periphery (data not shown). In summary, the designed proteins CCTO and CCEO appear to form folded, active RNases H.

Free Energy of Unfolding of CCEO and CCTO at 25 °C. The stabilities of CCTO and CCEO were measured by monitoring the GdmCl- and thermal-induced unfolding by following the CD signal at 222 nm (Figure 2B). Both proteins unfold cooperatively and behave in an apparent two-state manner; the $\Delta G°_{unf}$ for each denaturant melt was fit using a two-state model with a linear dependence on denaturant (*18*, *27*). GdmCl-induced denaturation of CCTO gives a stability of 9.6 ± 1.0 kcal/mol (25 °C) and a denaturant dependence or *m*-value of 4.2 ± 0.4 kcal/(mol·M); it is more stable than CtepRNH*, which contributed the core region, but less stable than periphery contributor TthRNH*. CCEO, surprisingly, is more stable under chemical denaturation than either of its parent proteins, CtepRNH* or EcoRNH*, with a $\Delta G°_{unf}$ of 9.8 ± 0.6 kcal/mol (25 °C) and *m*-value of 5.1 ± 0.3 kcal/(mol·M).

Thermal denaturation of CCTO yields a T_m of 69.4 ± 0.6 °C, and, like both of its parent proteins, thermal denaturation of CCTO is completely reversible after gradual heating to 95 °C and rapid cooling to 25 °C. Thermal denaturation of CCEO yields a T_m of 65 ± 0.5 °C, extremely similar to those of its parents. Rapid cooling from 95 to 25 °C was approximately 90% reversible, unlike EcoRNH*, which under similar conditions is completely irreversible. Thus, the folding cores of these RNases H appear to correlate with the reversibility of the protein's folding when thermally denatured.

Thermodynamic Analyses of Chimeric Proteins. The $\Delta G°_{unf}$ of both CCEO and CCTO was determined at a variety of temperatures to create a so-called protein stability curve. When fit independently, the *m*-values displayed a temperature-dependent variation which fit well to a linear model, and therefore we adopted the procedure described by Nicholson et al. (*25*), which accounts for a such a temperature-dependent *m*-value (see Experimental Procedures). The resulting stability curves ($\Delta G°_{unf}$ vs temperature) were fit using the Gibbs−Helmholz equation and are shown in Figure 3 together with the published results for the three parent proteins (*E. coli*, *T. thermophilus*, and *C. tepidum* RNases H*). The CCTO chimera is less stable than TthRNH* and more stable than CtepRNH* at all temperatures. Additionally, CCTO maintains the low $\Delta C°_P$ associated with the two parent thermophilic proteins (1.7 ± 0.1 kcal/(mol·K)). The CCEO chimera, however, shows a greater stability than both of its parents, EcoRNH* and CtepRNH*, at all measured temperatures even though its melting temperature is

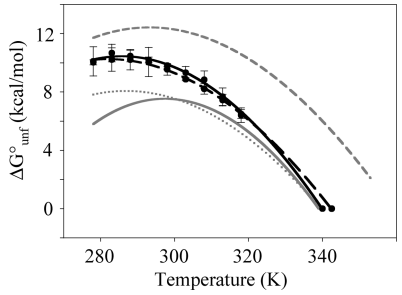


FIGURE 3: Protein stability curves of CCEO (solid black line and triangles) and CCTO (dashed black line and circles) overlaid on the stability curves of *E. coli* (solid gray line), *C. tepidum* (dotted gray line), and *T. thermophilus* (dashed gray line) RNases H*. Each point represents the average of two to three $\Delta G°_{unf}$s determined from isothermal GdmCl-induced denaturation experiments, with one additional point for each T_m, determined by reversible thermal denaturation. Lines represent fits to the Gibbs−Helmholz equation. Data were taken in 50 mM KCl, 20 mM NaOAc, pH 5.5, with varying concentrations of GdmCl at varying temperatures. *E. coli* and *T. thermophilus* RNase H* stability curves reproduced from ref *11*. CtepRNH* stability curve reproduced from ref *9*.

similar to both. The $\Delta C°_P$ of CCEO (2.1 ± 0.2 kcal/(mol·K)) is between those of CtepRNH* and EcoRNH*, something that was not seen with the previous chimeras studied (*15*). Both chimeric proteins preserve a temperature of maximum stability similar to that of CtepRNH* (13 °C for CCEO, 8 °C for CCTO, and 14 °C for CtepRNH*, as compared to 25 and 20 °C for EcoRNH* and TthRNH*, respectively). All thermodynamic parameters determined from this stability curve analysis are listed in Table 1.

Initial Characterization and Thermodynamics of C. tepidum RNase H I53D.* Since the chimeras constructed here both display the unusually low $\Delta C°_P$ associated with the thermophilic RNases H, we decided to probe the role of the core region further by site-specific mutagenesis. In TthRNH*, the mutation I53D appears to disrupt residual structure in the unfolded state, resulting in an increase in $\Delta C°_P$ (*16*). To test if the homologous residue also plays a similar role in CtepRNH*, we created the homologous variant (called CtepRNH* I53D). Both CD spectra and activity assays demonstrate that CtepRNH* I53D adopts the RNase H-like fold (data not shown). Thermodynamic analyses, analogous to those described above, demonstrate that, like in TthRNH*, the mutation I53D severely destabilizes CtepRNH* ($\Delta G°_{unf}$ of CtepRNH* I53D is 4.1 ± 0.3 kcal/mol at 25 °C). In contrast to the TthRNH* results, however, introducing the I53D mutation to CtepRNH* does not appear to affect the $\Delta C°_P$ of the protein ($\Delta C°_P$ of CtepRNH* I53D is 1.7 ± 0.2 kcal/(mol·K)) (Figure 4A).

Design, Expression, and Characterization of C. tepidum RNase H variants at position 56*. While the chimera studies suggest that the core of RNase H plays a critical role in the unusual ΔC°_P of CtepRNH*, the failure of the I53D mutation to alter the ΔC°_P of the protein, along with the intermediate ΔC°_P value for the CCEO variant, prompted us to examine the folding core more closely. Investigation of the hydrophobic core of the protein using a program that selectively identifies ILV clusters (*28*) suggested that L56 is likely to be a more central residue to the hydrophobic network in the folding core of CtepRNH* (Figure 5). If this ILV-based hydrophobic network in the core of the protein is crucial to the residual structure in the unfolded state, mutations at residue 56 may have a more notable effect. To test this, two variants of CtepRNH*, L56A and L56D, were constructed, purified, and analyzed. Initial CD and activity studies showed that L56A is well folded and active under standard RNase H* conditions. L56D, however, was poorly folded, as evidenced by little secondary structure by CD, the lack of a cooperative transition in a denaturant melt, and almost no activity. Apparently, the insertion of this potentially charged residue was so deleterious that the protein could not fold. Therefore, we created a potentially less destabilizing yet still hydrophobic-to-polar mutation, L56S. L56S is folded and active under standard conditions and, thus, presented as a good candidate for thermodynamic studies.

Table 1: Thermodynamic Parameters of *E. coli*, *C. tepidum*, *T. thermophilus*, Chimeras, and Various Mutant RNases H*[a]

	ΔG°_{unf} (kcal/mol) (25 °C)	ΔC°_P [kcal/(mol·K)]	T_m (°C)
EcoRNH*	−7.5 ± 0.2	2.7 ± 0.2	67
TthRNH*	−12.2 ± 0.4	1.8 ± 0.1	86
CtepRNH*	−8.1 ± 0.5	1.7 ± 0.1	66.5
CCEO	−9.8 ± 0.6	2.1 ± 0.2	65
TCEO	−7.5 ± 0.5	1.6 ± 0.2	76
CCTO	−9.6 ± 1.0	1.7 ± 0.1	69.5
ECTO	−5.6 ± 0.3	2.4 ± 0.3	61
CtepRNH* I53D	−4.1 ± 0.3	1.7 ± 0.2	50.5
CtepRNH* L56A	−2.8 ± 0.3	1.2 ± 1.1	47
CtepRNH* L56S	−1.1 ± 0.4	2.5 ± 0.9	33.5
TthRNH* I53D	−7.0 ± 0.5	2.4 ± 0.2	69
TthRNH* L56S	−7.0 ± 0.4	2.65 ± 0.5	56.5

[a]Data for *E. coli* and *T. thermophilus* RNases H* taken from ref *11*, data for TCEO and ECTO taken from ref *15*, and data for *T. thermophilus* I53D taken from ref *16*. Errors reported in ΔC°_P for *C. tepidum* and its variants were calculated as described in Experimental Procedures.

Thermodynamic Analysis of CtepRNH L56A and L56S*. Mutation of residue 56 to either alanine or serine was extremely destabilizing. ΔG°_{unf} (25 °C) for CtepRNH* L56A is 2.75 ± 0.3 kcal/mol and 1.10 ± 0.4 kcal/mol for CtepRNH* L56S, which translates into $\Delta\Delta G^\circ_{unf}$s relative to CtepRNH* of 5.35 and 7.00 kcal/mol, respectively. Like CtepRNH*, both mutants are maximally stable near 10 °C (11.5 °C for L56S and 5.9 °C for L56A). As predicted, the more conservative mutation, L56A, does not alter the ΔC°_P of CtepRNH*, maintaining a low value of 1.2 ± 1.1 kcal/(mol·K). On the other hand, the polar mutation, L56S, dramatically affects the ΔC°_P, raising it to 2.5 ± 0.9 kcal/(mol·K) (Figure 4B). The stability curves of these proteins are not as well determined as others in this study. The low stability and correspondingly low T_ms of each protein

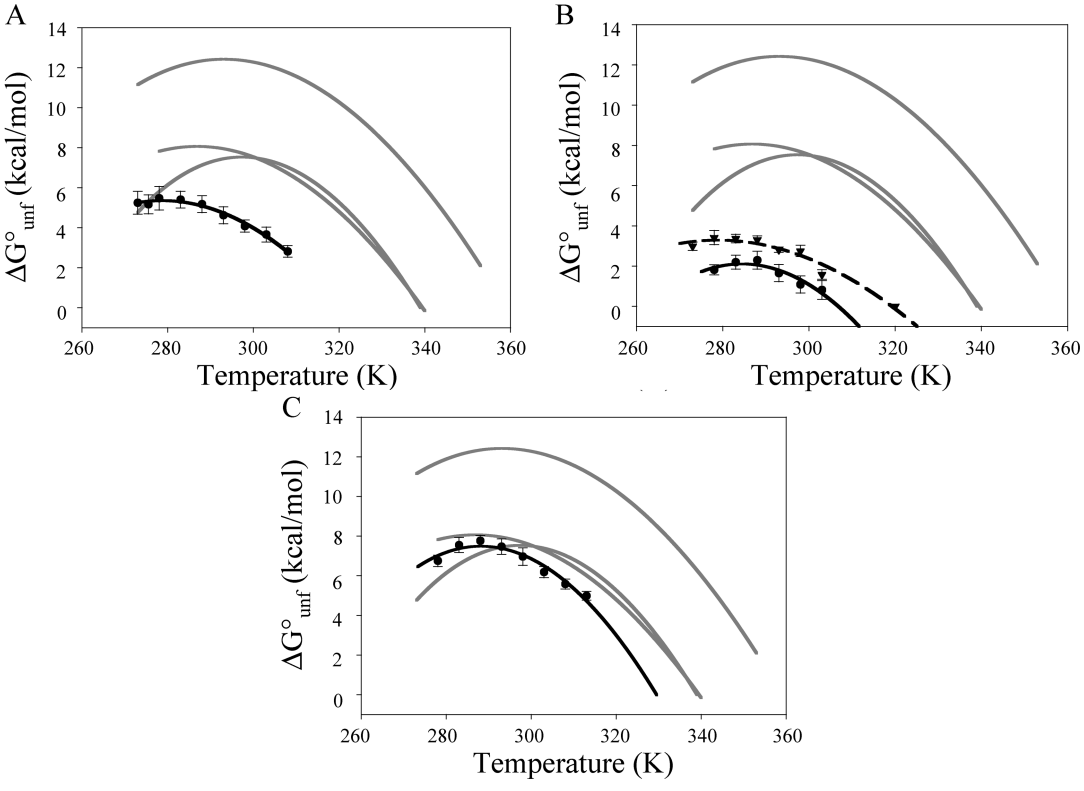


FIGURE 4: Protein stability curves of various mutant RNases H* overlaid on the stability curves of *E. coli*, *C. tepidum*, and *T. thermophilus* (gray lines) RNases H*. (A) CtepRNH* I53D in black circles and solid black line. (B) CtepRNH* L56A (dashed black line and triangles) and L56S (solid black line and circles). (C) TthRNH* L56S in solid black line and circles. Each point represents the average of two to three ΔG°_{unf}s determined from isothermal GdmCl-induced denaturation experiments. Lines represent fits to the Gibbs−Helmholz equation. Data were taken in 50 mM KCl, 20 mM NaOAc, pH 5.5, with varying concentrations of GdmCl at varying temperatures. *E. coli* and *T. thermophilus* RNase H* stability curves reproduced from ref *11*. CtepRNH* stability curve reproduced from ref *9*.

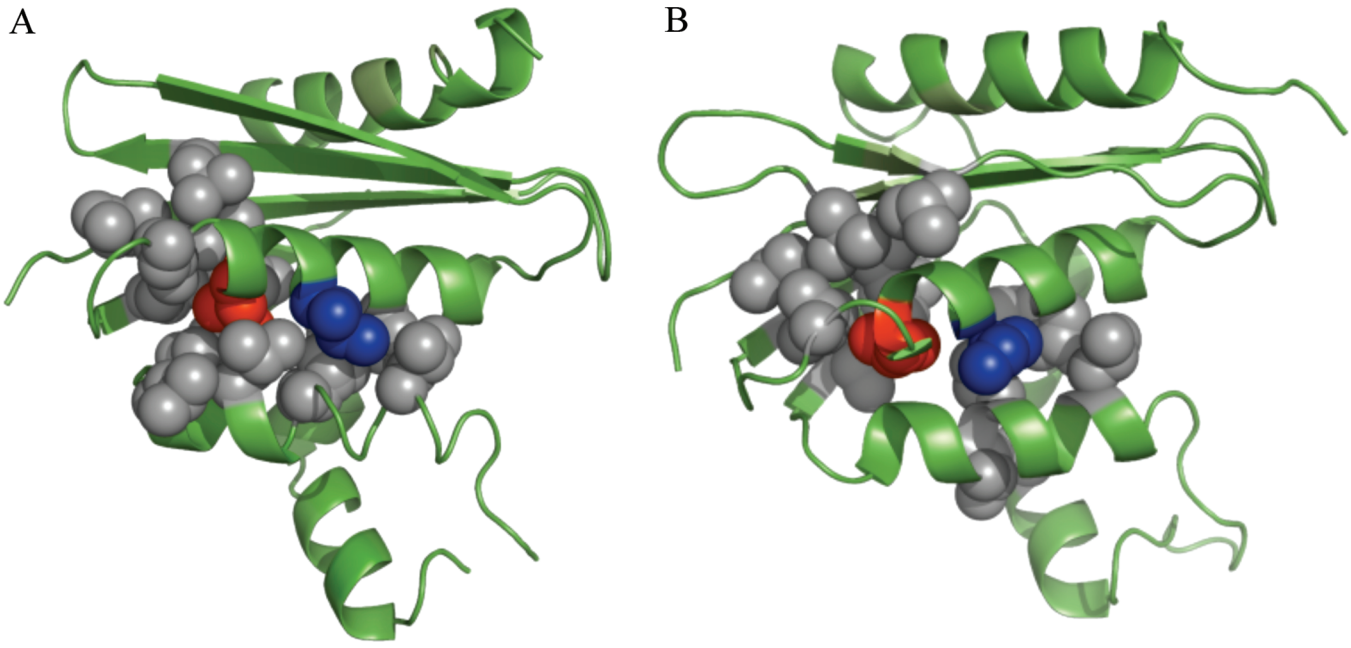


FIGURE 5: Structures depicting the positions of the ILV clusters in CtepRNH* (A) and TthRNH* (B). Gray spheres represent ILV residues, red spheres represent residue L56, and blue spheres represent residue I53.

limit the range of temperatures at which the ΔG°_{unf} can be measured and contribute to higher errors in the fit. Despite some uncertainty in the data, the trend is clear: CtepRNH* L56S has an increased ΔC°_P while CtepRNH* I53D and CtepRNH* L56A do not.

Investigation of Residue 56 in T. thermophilus RNase H. The mutations at position 56 in CtepRNH* suggest that this site (like I53 in TthRNH*) may play an important role in the unusually low ΔC°_P of this protein. The question then arises whether this residue plays an important role in the residual structure in the unfolded state of TthRNH* as well. In other words, is the residual structure in the unfolded state of CtepRNH* a different set of residues or just more restricted? To investigate this, we turned back to TthRNH* and interrogated the structurally homologous position, L56. TthRNH* L56S is well folded with a cooperative, apparent two-state transition under both temperature and chemical denaturation. The ΔG°_{unf} for TthRNH* L56S is 7.0 ± 0.4 kcal/mol (25 °C). While the protein is significantly destabilized, the extremely high stability of the parent protein, TthRNH*, allows for a well-determined stability curve, and the ΔC°_P of this protein is 2.65 ± 0.5 kcal/(mol·K) (Figure 4C), substantially higher than its parent protein TthRNH* (1.8 kcal/(mol·K)). Thus, residue 56 in CtepRNH* also plays an important role in the residual structure in the unfolded state of TthRNH*.

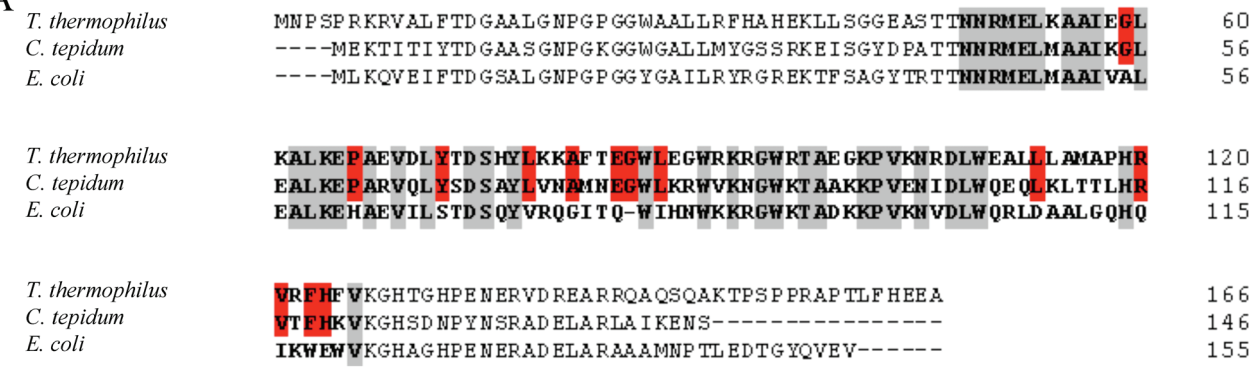
DISCUSSION

In this study, we probed the RNase H from the moderately thermophilic organism *C. tepidum* (CtepRNH*) to investigate how the unusually low ΔC°_P is encoded in the sequence and structure of thermophilic RNases H. Like the RNase H from *T. thermophilus*, CtepRNH* displays an unusually low ΔC°_P (the measured ΔC°_P is lower than that expected based on the burial of nonpolar surface area in the native state, calculated to be between 2.1 and 2.4 kcal/(mol·K) (*9*, *11*, *18*)). These low ΔC°_Ps result in a broadened protein stability curve and therefore contribute to the increased stability at higher temperatures. By analogy to previous comparative studies with RNases H from *T. thermophilus* and *E. coli*, this low ΔC°_P was assumed to arise from residual structure in the unfolded state and to be an important feature in the protein's ability to function at higher temperatures. Studies on the homologues from *E. coli* and *T. thermophilus* identified the folding core as the region responsible for determining the protein's ΔC°_P; chimeras between the core and periphery of each protein resulted in functional RNases H with a ΔC°_P corresponding to the parent contributing the core (*15*). Previously, we showed that RNase H from *C. tepidum* also folds through an early intermediate (*9*), which we believe to be composed of the same core region.

To confirm that the presumed early folding region of *C. tepidum* RNase H is also responsible for the low ΔC°_P of this protein, we made two chimeras which contain the CtepRNH* folding core and either the *E. coli* or *T. thermophilus* periphery. Remarkably, both chimeras formed active and stable RNases H. This ability to mix and match regions of a protein is common when working with multidomain proteins that have separate structural and functional units, such as the SH3 domain or pdz domains. RNase H, however, is a small, apparently single domain protein. Undoubtedly, our ability to mix and match the core and periphery of these RNases H is related to the fact this core region constitutes the early folding intermediate of these proteins (*10*, *22*). Rapid structure formation in the core implies that this part of the protein forms a somewhat autonomous unit, encoding its own structure. In fact, studies of isolated RNase H fragments from both *E. coli* and *T. thermophilus* have shown that this region can fold independently (*17*, *29*). Our success with the homologous folding core of CtepRNH* clearly supports our hypothesis that these residues also form the early folding region of this protein and suggests that folding cores of other proteins may well be suitable candidates for modulation and transplantation as a unit.

The ΔG°_{unf} of CCEO was found to be higher than those of EcoRNH* and CtepRNH* at all temperatures. This higher stability may arise from the increased stability of the core of CtepRNH* relative to that of EcoRNH*. Previous data have shown that the ΔG°_{unf} of the core of CtepRNH* is 4.1 kcal/mol,

A



B

	T. thermophilus	*C. tepidum*
C. tepidum	57% / 62%	
E. coli	57% / 55%	53% / 54%

FIGURE 6: Sequence alignment of RNases H. (A) Alignment of TthRNH*, CtepRNH*, and EcoRNH*. Residues in bold are in the core region, residues highlighted in gray are core residues that are identical in all three proteins, and residues highlighted in red are core residues that are identical in TthRNH* and CtepRNH* but differ in EcoRNH*. (B) Percent sequence identities between TthRNH*, CtepRNH*, and EcoRNH*. In all cases, the first number listed is the percent identity of the whole protein sequence and the second number is the percent identity of the core region.

while that of EcoRNH* is 3.55 kcal/mol (*9*, *22*). Both native proteins have approximately the same stability, so the core of CtepRNH* must account for a higher percentage of the overall stability of the protein than that of EcoRNH*. It appears that in the CCEO chimera we have combined the more stable core of CtepRNH* with the relatively more stable periphery of EcoRNH*. This combination has resulted in a protein with a stability increased from that of either parent protein, despite the creation of a new, nonoptimized interfacial region. The ΔC°_P of CCEO, between that of EcoRNH* and CtepRNH*, both closer to that of CtepRNH* and lower than predicted for a protein of this size, indicates that the ΔC°_P is mostly governed by the folding core in this family of proteins (Table 1 and Figure 3).

The ΔC°_P for both *C. tepidum* RNase H chimeras created here are significantly lower than that expected based on the assumed native state structure (~2.7 kcal/(mol·K)). These results agree with the previous studies on chimeras between *E. coli* and *T. thermophilus* RNases H and support the hypothesis that this core region is the primary determinant of the residual unfolded state structure. CCTO has an identical ΔC°_P to both its parent proteins, which implies that either the CtepRNH* core completely dictates this feature or any extension outside the core region is similar in the two parent proteins. The latter is not unreasonable as both proteins arise from thermophilic organisms. The slightly higher ΔC°_P observed for CCEO may suggest that our imposed boundaries do not completely encompass the residues required for this property, and hence some of the residues encoding the residual structure reside either at the interface or outside of the bounded region defined as the core.

Since the cores of the thermophilic proteins track with the unusually low ΔC°_P, we assume that this region of TthRNH* and CtepRNH* shares properties with each other that are not shared with the mesophilic homologue. It is interesting to note, therefore, that while the pairwise identities between the three proteins are all between 53% and 57% (Figure 6), the presumed cores of TthRNH* and CtepRNH* are 62% identical while the pairwise identity between either of the thermophiles and the mesophilic protein is 54−55%. Out of the ~80 residues that make up the presumed core, almost half (36) are conserved in all three proteins. While some of these 36 residues are likely to be involved in defining this low ΔC°_P, they cannot be the only ones responsible since we observed that mutation of completely conserved sites such as I53 or L56 has dramatic effects on ΔC°_P. It is more likely that other core residues in the mesophilic protein are responsible for disrupting this hydrophobic network that makes up the residual structure. Interestingly, 13 of the residues in the core that are not conserved between all three proteins are identical in just the two thermophilic homologues. As the set of RNases H that encode this unusually low ΔC°_P enlarges, a rigorous bioinformatic analysis may lend insight into the genetic coding of this unusual feature.

Using our site-specific mutagenesis approach, we have begun to identify the residues that contribute to the low ΔC°_P and, presumably, therefore also contribute to the residual structure in the unfolded state. While any mutation of a hydrophobic core residue is likely to affect the stability of the protein, only variants that disrupt the unfolded state structure will result in a restoration of the higher ΔC°_P seen in the mesophilic protein. Previous studies on TthRNH* and EcoRNH* made point mutants at residue I53, a completely buried residue located in the center of the A helix. While the variant I53D destabilizes both EcoRNH* and TthRNH*, it also selectively alters the ΔC°_P of TthRNH*, restoring it to near the expected mesophilic value. Similarly, our results here with the analogous mutation in CtepRNH* demonstrated a large destabilizing effect. Surprisingly, however, introducing this dramatic mutation left the ΔC°_P completely unchanged. This unexpected result suggests that the residues making up the hydrophobic network and therefore residual structure in the unfolded state must be somewhat different in the two proteins.

Matthews and co-workers have suggested that clusters of the hydrophobic residues isoleucine, leucine, and valine (ILV clusters) play an important role in early collapse during protein folding (*28*), so we hypothesized that these ILV clusters could be related to the residual structure in the unfolded state. By identifying changes between the ILV networks found in these two proteins, we attempted to uncover differences in the composition of structure in the unfolded state. Indeed, such an analysis on the folding cores of TthRNH* and CtepRNH* revealed that TthRNH* seems to have two main ILV clusters, with I53 central to one cluster, while CtepRNH* has one larger cluster that does not involve I53 as a central residue (Figure 5). Furthermore, the residues involved in the CtepRNH* cluster extend past the boundaries of the folding core as defined in previous studies, potentially explaining the observed intermediate ΔC°_P for CCEO.

Our mutagenesis studies of residue 56 suggest that this site has profound consequences on the protein. The mutation L56D appears so disruptive that the native protein is too unstable for biophysical analysis. Introduction of a polar residue (L56S) raised the ΔC°_P to create a CtepRNH* with apparent "mesophile-like" ΔC°_P. Thus, by analogy, residue 56 appears be a key component to the residual structure in the unfolded state of CtepRNH*. A similar analysis of this position in TthRNH* suggests that it also plays an essential role in the folding of TthRNH*. Thus, we believe that the residual structure in TthRNH* involves both ILV clusters identified, while the residual structure in CtepRNH* is limited to one main cluster that is unaffected by mutations to residue 53. These data demonstrate that the hydrophobic clustering in the unfolded state of CtepRNH* has different boundaries than that of TthRNH. The specific interactions that create and define the low ΔC°_P in the thermophilic RNases H are not conserved between the two.

Proteins that function at high temperature employ various strategies to fold and function at the required temperature. RNase H is one example of a large class of enzymes that achieve this in part through a broadening of the protein stability curve (i.e., a lowering of ΔC°_P). Our results suggest that encoding residual structure in the unfolded state of the protein in the region of the folding core may be a general strategy used by the RNase H family of enzymes. It will be interesting to see if other proteins that employ this change in ΔC°_P strategy do so through the unfolded state structure in the region of the protein that folds first. While the general strategy of a low ΔC°_P appears conserved, our results show that at the residue level the details of this strategy are protein specific. Through our protein engineering approach, we have successfully decoupled stability and residual structure in the unfolded state, leading to the creation of a more "thermophile-like" protein as defined through a raised overall stability and lowered ΔC°_P (CCEO). This approach should have important implications for the design of novel enzymes that function at elevated temperatures.

ACKNOWLEDGMENT

We thank Jesse Dill for help with data analysis, C. R. Matthews and Sagar Kathuria for help with the ILV cluster analysis, and the entire Marqusee laboratory for helpful discussion.

SUPPORTING INFORMATION AVAILABLE

Examples of standard RNase H activity assays on the RNase H variants CCEO, CCTO, CtepRNH* L56A, and CtepRNH* L56D and raw data from sample denaturant melts used to create the reported protein stability curves. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H., and Morikawa, K. (1993) Crystal structure of ribonuclease H from *Thermus thermophilus* HB8 refined at 2.8 Å resolution. *J. Mol. Biol. 230*, 529–542.
2. Goedken, E. R., Keck, J. L., Berger, J. M., and Marqusee, S. (2000) Divalent metal cofactor binding in the kinetic folding trajectory of *Escherichia coli* ribonuclease HI. *Protein Sci. 9*, 1914–1921.
3. Knapp, S., Kardinahl, S., Hellgren, N., Tibbelin, G., Schafer, G., and Ladenstein, R. (1999) Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon *Sulfolobus acidocaldarius* at 2.2 Å resolution. *J. Mol. Biol. 285*, 689–702.
4. Knapp, S., de Vos, W. M., Rice, D., and Ladenstein, R. (1997) Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. *J. Mol. Biol. 267*, 916–932.
5. Korndorfer, I., Steipe, B., Huber, R., Tomschy, A., and Jaenicke, R. (1995) The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol. 246*, 511–521.
6. Szilagyi, A., and Zavodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure 8*, 493–504.
7. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403–410.
8. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1992) Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol. 223*, 1029–1052.
9. Ratcliff, K., Corn, J., and Marqusee, S. (2009) Structure, stability, and folding of ribonuclease H1 from the moderately thermophilic *Chlorobium tepidum*: comparison with thermophilic and mesophilic homologues. *Biochemistry 48*, 5890–5898.
10. Hollien, J., and Marqusee, S. (2002) Comparison of the folding processes of *T. thermophilus* and *E. coli* ribonucleases H. *J. Mol. Biol. 316*, 327–340.
11. Hollien, J., and Marqusee, S. (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry 38*, 3831–3836.
12. Hollien, J., and Marqusee, S. (1999) Structural distribution of stability in a thermophilic enzyme. *Proc. Natl. Acad. Sci. U.S.A. 96*, 13674–13678.
13. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Biol. 3*, 782–787.
14. Fischer, K. F., and Marqusee, S. (2000) A rapid test for identification of autonomous folding units in proteins. *J. Mol. Biol. 302*, 701–712.
15. Robic, S., Berger, J. M., and Marqusee, S. (2002) Contributions of folding cores to the thermostabilities of two ribonucleases H. *Protein Sci. 11*, 381–389.
16. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M., and Marqusee, S. (2003) Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl. Acad. Sci. U.S.A. 100*, 11345–11349.
17. Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M., and Marqusee, A. S. (1999) Folding of an isolated ribonuclease H core fragment. *Protein Sci. 8*, 2251–2257.
18. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci. 4*, 2138–2148.
19. Gomez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins. *Proteins 22*, 404–412.
20. Eisen, J. A., Nelson, K. E., Paulsen, I. T., Heidelberg, J. F., Wu, M., Dodson, R. J., Deboy, R., Gwinn, M. L., Nelson, W. C., Haft, D. H., Hickey, E. K., Peterson, J. D., Durkin, A. S., Kolonay, J. L., Yang, F., Holt, I., Umayam, L. A., Mason, T., Brenner, M., Shea, T. P., Parksey, D., Nierman, W. C., Feldblyum, T. V., Hansen, C. L., Craven, M. B., Radune, D., Vamathevan, J., Khouri, H., White, O., Gruber, T. M., Ketchum, K. A., Venter, J. C., Tettelin, H., Bryant, D. A., and Fraser, C. M. (2002) The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc. Natl. Acad. Sci. U.S.A. 99*, 9509–9514.
21. Doering, D. S., and Matsudaira, P. (1996) Cysteine scanning mutagenesis at 40 of 76 positions in villin headpiece maps the F-actin

binding site and structural features of the domain. *Biochemistry 35*, 12677–12685.
22. Raschke, T. M., Kho, J., and Marqusee, S. (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat. Struct. Biol. 6*, 825–831.
23. Santoro, M. M., and Bolen, D. W. (1992) A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry 31*, 4901–4907.
24. Becktel, W. J., and Schellman, J. A. (1987) Protein stability curves. *Biopolymers 26*, 1859–1877.
25. Nicholson, E. M., and Scholtz, J. M. (1996) Conformational stability of the *Escherichia coli* HPr protein: test of the linear extrapolation method and a thermodynamic characterization of cold denaturation. *Biochemistry 35*, 11369–11378.
26. Black, C. B. a. C., and J.A. (1994) Magnesium activation of ribonuclease H—evidence for one catalytic metal ion. *Inorg. Chem. 33*, 5805–5808.
27. Pace, C. N., and Laurents, D. V. (1989) A new method for determining the heat capacity change for protein folding. *Biochemistry 28*, 2520–2525.
28. Wu, Y., Vadrevu, R., Kathuria, S., Yang, X., and Matthews, C. R. (2007) A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the alpha subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol. 366*, 1624–1638.
29. Zhou, Z., Feng, H., Ghirlando, R., and Bai, Y. (2008) The high-resolution NMR structure of the early folding intermediate of the *Thermus thermophilus* ribonuclease H. *J. Mol. Biol. 384*, 531–539.