Title
Beyond the native state: Exploring the role of partially folded conformations on the protein energy landscape

Author
Connell, Katelyn Blair

Publication Date
2010

Peer reviewed|Thesis/dissertation

Beyond the native state: Exploring the role of partially folded conformations on the protein energy landscape

By

Katelyn Blair Connell

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Susan Marqusee, Co-Chair
Professor Judith Klinman, Co-Chair
Professor David Wemmer
Professor Harvey Blanch

Fall 2010

Abstract

Beyond the native state: Exploring the role of partially folded conformations on the protein energy landscape

by

Katelyn Blair Connell

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Susan Marqusee, Chair

Proteins can sample a variety of partially folded conformations during the transition between the unfolded and native states. The role of such intermediates is a matter of considerable debate, but it is clear that characterization of these partially folded species is crucial for understanding protein folding and function. A single amino acid change can convert *E. coli* ribonuclease H from a three-state folder that populates a kinetic intermediate to one that folds in an apparent two-state fashion. We have compared the folding trajectories of the three-state and two-state RNases H, proteins with the same native state topology but altered regional stability, using a protein engineering approach. Our data indicate that that both versions of RNase H fold through a similar trajectory with similar high-energy conformations. This suggests that formation of specific partially folded conformations may be a general feature of protein folding that can promote, rather than hinder, efficient folding.

To better understand the robust role this high-energy species plays in folding, we set out to trap the transient intermediate of RNase H at equilibrium by selectively destabilizing the region of the protein known to be unfolded in this species. We find that the intermediate is undetectable in a series of HSQC's, revealing the dynamic nature of this partially folded form on the timescale of NMR detection. This result is in contrast to studies in which the structures of trapped intermediates are solved by NMR, indicating that the they are well-packed and native-like. The dynamic nature of the RNase H intermediate may be important for its role as an on-pathway, productive species that promotes efficient folding. An analogous intermediate is populated on the kinetic trajectory of RNase H from *T. thermophilus*, an organism that grows optimally at a temperature 30 ^{o}C higher than *E. coli*. To understand how two proteins that share identical structures can function in such different environments, we looked for differences in their energetics by comparing equilibrium mimics of their high-energy intermediates. We find potential differences in the dynamic properties of the intermediates, which may provide insight into how proteins with the same native structure can exhibit vastly different biophysical behavior.

In contrast to globular proteins such as RNase H, repeat proteins are tandem arrays of repeating structural units that have no long-range contacts. In these modular

domains, the majority of native contacts could be maintained in the face of partial unfolding. Repeat proteins therefore offer a unique architecture for exploring the extent of cooperativity and roughness on the energy landscape. To understand how a modular system builds cooperativity into its energetics, and to explore the origins and limits of this cooperativity, we studied the behavior of the Notch ankyrin domain in the optical tweezers, a single molecule mechanical tool. The forced unfolding of the Notch ankyrin domain occurs in one or two steps when manipulated in the optical tweezers. Though the unfolding pathway is heterogenous compared to that observed in bulk studies, there is a limit to the degree of uncoupling of individual repeats. We compare these results to the unfolding behavior of the Notch ankyrin domain in the atomic force microscope obtained by our collaborators for this project. This offers some insight into the apparent difference in solution and AFM unfolding of ankyrin repeat proteins.

For my parents

List of Tables

Chapter II.

Chapter III.

Chapter IV.

List of Figures

Chapter I.

Chapter II.

Chapter III.

Chapter IV.

Chapter V.

Acknowledgements

My time in graduate school has transformed me both professionally and personally, and I have many people to thank for their impact on me during this time. First, I thank my advisor, Susan Marqusee, for her key role in my scientific education. With her guidance I feel I have grown tremendously as a scientist. Her intellect, support, and kindness are greatly appreciated. I would like to thank all of the members of the Marqusee lab for creating a friendly and supportive environment. I cannot imagine having a better group of people to grow, laugh, and collaborate with; I will truly miss working with all of you. In particular, I would like to thank Katie Tripp, whose arrival to the Marqusee lab coincided with my own. Her friendship has been a daily source of wisdom, comfort, and happiness.

My family has also played a key role in my academic success. I cannot begin to express my depth of gratitude for my parents, who have provided me with the foundation to succeed both professionally and personally. None of this is possible without you. I thank my brothers, Bryan and Kevin, and their families, for love and support and for adding so much to my life. A very heartfelt thank you to my mother- and father-in-law, who have been a constant source of encouragement, advice, and love. Finally, thank you to my husband, Weston. I have learned more from you than could be explained in a thesis, and I would not be the same without your love and understanding.

Chapter I

Introduction to
Beyond the native state: Exploring the role of partially folded conformations on the protein energy landscape

1.1 PROTEIN ENERGY LANDSCAPES

Proteins are essential parts of organisms that participate in virtually every cellular process. Their vital roles include enzyme catalysis, maintenance of structural integrity, and mediating signaling. In order to carry out its specific function, a protein polymer must assume a specific conformation. Protein folding is the fascinating biological process by which the linear chain of amino acids adopts a unique three-dimensional structure, and it remains one of the mysteries of molecular biology. All of the information necessary for three-dimensional structure formation is encoded by the order of the amino acids[1], but we cannot yet predict the native fold given only the primary sequence. Furthermore, a vast conformational space is available to the protein chain, yet proteins typically fold in seconds or less[2]. What physical laws and interactions govern protein folding pathways?

A useful conceptualization of this fundamental question is the protein energy landscape, an example of which is shown in Figure 1[3; 4]. The primary sequence encodes not only the final native structure but also this entire landscape, which describes all the conformations accessible to a protein and the barriers between them. In this visualization, the horizontal axis represents the conformation of the protein, and the vertical axis represents the internal energy of that conformation. The energy of a given conformation will change depending on the conditions of the system. This figure represents folding conditions, and the native state occupies the minimum at the bottom of the landscape, with a large energetic separation between it and any partially or fully unfolded forms. In order to study the important features of the energy landscape, it must be perturbed in some way. The populations of and barriers between the different species are typically manipulated by factors such as chemical denaturant and temperature.

The energy landscape is important for addressing questions beyond those related to the folding process; fluctuations from the native state are necessary for function. Proteins must undergo conformational changes for allosteric regulation and in order to catalyze reactions, for instance. In fact, an increasing number of intrinsically disordered proteins, which fold upon finding a binding partner, are being identified[5]. To fully grasp how an amazing diversity of life has evolved, it is critical to consider that the entire energy landscape, not just the native state of a protein, is subjected to the forces of evolution.

The work in this thesis focuses on non-native conformations known as intermediates and seeks to understand how they fit into the overall folding and function of a protein. Intermediates are partially folded species that occupy a local energy minimum on the landscape. This well is broader than that of the native state, indicating that it is a more heterogeneous ensemble of conformers. Intermediates offer a wealth of information regarding how conformational space is narrowed as a protein folds. Unfortunately, they are difficult to study directly. Many small proteins are observed to fold directly to the native state without the buildup of any such intermediates[6]. This two-state, cooperative folding transition makes it challenging to study the mechanistic details of structure formation. When a protein does fold through an observable intermediate state on its folding trajectory, it is populated only transiently. At equilibrium, the partially folded states represent a tiny fraction of the overall population due to a large energetic separation from the native state. It is important to note that a partially folded conformation accessible to a protein at equilibrium is not necessarily sampled on its

kinetic trajectory[7]; however, in some cases, such as for apomyoglobin[8], cytochrome c[9], and RNase H[10], intermediates detected by kinetic and equilibrium experiments are related[11]. The work presented here primarily addresses kinetic intermediates, and in the following section, some important experimental techniques for detecting and characterizing such intermediates are discussed.

1.2 CHARACTERIZING KINETIC INTERMEDIATES

Chevron Plots

The first test for determining if a protein populates an intermediate on its folding pathway is to analyze the denaturant dependence of the folding and unfolding kinetics[12]. These experiments are typically performed using a stopped-flow mixing device, in which concentrated unfolded protein is rapidly diluted into folding conditions, in the case of refolding experiments. The resulting signal change can be monitored by a variety of spectroscopic techniques. There is a mixing time typically on the order of milliseconds associated with the stopped-flow mixer, during which reliable signal cannot be obtained. Continuous flow mixers are capable of microsecond mixing times, but these are currently being developed and less commonly used. A graph of the natural log of the observed rate constant versus denaturant concentration is called a chevron plot. If a single barrier exists between unfolded and folded protein, the kinetic traces are described by a single exponential and the result will be a classic V-shaped chevron plot in which the denaturant dependence of the log of the rates is linear. This is indicative of two-state folding, a classic example of which is chymotrypsin inhibitor 2[13]. Simulated kinetic data for a two-state folding protein is shown in Figure 2a.

Deviations from linearity on either the folding or unfolding limb of the chevron often suggest the presence of a transient intermediate. One common type of refolding intermediate important for the work described in this thesis is called a burst phase intermediate[14]. This refers to a scenario in which much of the expected signal change occurs in the dead time of the stopped-flow mixing device followed by a much slower, observable signal change. The separation of time scales indicates that there are at least two processes occurring. The signal change observed to occur in the dead time may arise from the rearrangement of the denatured state in the new solvent conditions or the formation of a folding intermediate. If a partially folded intermediate is truly responsible, the burst phase signal will decrease sigmoidally with increasing denaturant concentration, indicative of a cooperative transition. This, combined with rollover in the folding limb of the chevron, strongly suggests the presence of a burst phase intermediate. Because the event occurring in the dead time is much faster than the observable step, a pre-equilibrium between the unfolded and intermediate states can be assumed. Rollover is observed because the fraction of intermediate that exists under the final folding conditions changes as a function of denaturant, and the observed rate (k_f) is proportional to that fraction (f_I) and the rate of folding from the intermediate to the native state (k_{IN}) under those conditions:

$$k_f = f_I k_{IN}$$

At low denaturant concentrations where the fraction of intermediate is high, the observed rate depends primarily on k_{IN}. As denaturant concentration increases, and the fraction of intermediate that is populated under these conditions decreases, the observed rate is proportional to product of K_{UI} and k_{IN}. Simulated data for a protein folding through a

burst phase intermediate is shown in Figure 2b. It is not possible to differentiate on-pathway from off-pathway intermediates from these data; additional experiments must be performed to address this question.

It is of course possible that folding intermediates elude detection because they exist after the rate-limiting transition state and hence are not significantly populated. In the case of an intermediate on the native side of the barrier, it might be expected that the intermediate could be detected in unfolding experiments. However, these are not usually observed because they are not stable in the high denaturant concentrations utilized in the unfolding experiments. Alternate methods have been employed to detect these unfolding intermediates. Intermediates on the native side of the barrier have been inferred from native state hydrogen exchange experiments for a redesigned apocytochrome b_{562}[15; 16], T4 lysozyme[16], and the third domain of PDZ from PSD-95[17], although, as discussed above, these are not necessarily correlated to the kinetic trajectory. NMR relaxation methods were used to detect an unfolding intermediate for acyl-coenzyme A binding protein[18].

When there is no direct evidence for the population of an intermediate, non-linearity in chevron plots can be interpreted as either a gradual movement of the transition state or a switch between two distinct transition states[19; 20]. The debates on this subject are heated and ongoing[21].

Phi-value analysis

A structural picture of folding intermediates provides valuable information about the factors that influence intermediate formation and folding pathways in general. Once it has been established that an intermediate is populated, it is possible to obtain information about what regions of the protein contribute to the structure of the intermediate. One method that probes tertiary structure formation in high-energy species on the folding trajectory, both intermediates and transition states, is phi-value (ϕ value) analysis[12; 22]. A ϕ value is the ratio of the affect of a mutation on the stability of the intermediate ($\Delta\Delta G_{UI}$) or transition state ($\Delta\Delta G^{\ddagger}$) to that of the native state (ΔG_{UN}):

$$\phi^{I} = \Delta\Delta G_{UI}/\Delta\Delta G_{UN} \qquad \phi^{\ddagger} = \Delta\Delta G^{\ddagger}/\Delta\Delta G_{UN}$$

When a mutation causes a change in global stability (the denominator of the above equations), this change must be reflected in either the folding or unfolding rate. The chevron plot analysis described above is carried out for the mutants and compared to the wild type protein. The stability of the intermediate can be determined from the fits of the burst phase amplitudes and the chevron. The change in the transition state stability is calculated by comparing the rates for folding and unfolding.

A ϕ value of 1 indicates that the mutation has the same energetic affect on the intermediate or transition state as it does on the native state, and we conclude that this residue must be important for stabilizing the intermediate or the transition state. A ϕ value of 0 indicates that the mutation has very little effect on the stability of the intermediate or transition state; thus, it must not be structured in the high-energy species in question. Intermediate ϕ values can be interpreted in two ways; they can indicate partial structure formation at the residue in question or that the residue contributes differently in different molecules.

Pulse-labeling hydrogen/deuterium exchange

Pulse-labeling hydrogen/deuterium exchange is a technique that probes secondary structure formation at various time points during refolding. Deuterated protein is refolded into protonated buffer, and a high pH pulse initiates hydrogen exchange with the

solvent. An amide deuterium is free to exchange with solvent protons when it is not structured; once that deuterium forms hydrogen bonds of stable secondary structure, it is prevented from exchange. A quench buffer halts exchange after the labeling period. Extent of exchange can be evaluated at the residue level by a ^{15}N-^{1}H HSQC NMR experiment, as protons are detectable but deuterons are not.

Equilibrium mimics of kinetic intermediates

There has been a great deal of interest in creating equilibrium populations of kinetic intermediates in order to gain access to high-resolution structural information. Two main approaches have been employed to achieve this goal, and both require knowledge of the structure of the intermediate obtained through ϕ value analysis, native state and pulse-labeling hydrogen exchange, or some combination of these methods. In one approach, the regions of the protein thought to be unfolded in the intermediate are removed at the genetic level, and the resulting fragment can be studied directly. This method was used to study the intermediate of *T. thermophilus* RNase H, and the NMR structure of the fragment revealed a well-packed species with residues that appear to represent a subset of those structured in native state[23]. This approach, however, does not take into account any role the unfolded regions of the intermediate might play in determining the dynamics and structure of the partially folded species. The second approach leaves the full-length protein intact, and the native state is selectively destabilized by introducing mutations into the regions believed to be unfolded in the intermediate. Bai and co-workers have employed this method extensively to study the intermediates of a redesigned apocytochrome b_{562}[24] andT4 lysozyme[25] at the residue level. It has also been used to characterize the misfolded intermediate of the immunity protein Im7[26; 27] and is the method described extensively in Chapters 3 and 4. The ability to study folding intermediates at atomic resolution can answer many fundamental questions about the nature of these elusive species.

1.3 THE ROLE OF INTERMEDIATES IN PROTEIN FOLDING

How ubiquitous is intermediate formation?

It is often written that proteins greater than 100 amino acids in length fold through one or more intermediate states; however the role and even the presence of intermediates on the folding trajectory are controversial subjects. Do intermediates comprise off-pathway species that hinder folding and lead to aggregation and disease, or do they promote efficient folding? Might they be ubiquitous, present even in the trajectories of proteins that appear to fold in a two-state manner?

The possibility that intermediate formation may be a general step in folding is highlighted in studies of families of proteins with a range of sequence similarities. In many of these cases, the general properties of the folding pathway are conserved while the details vary widely. The four-helix bacterial immunity proteins are one of the best-characterized examples. Im7 was observed to populate a kinetic intermediate composed of structure in helices I, II, and IV, while Im9, sharing 60% sequence identity, appeared two-state[28]. However, simply increasing hydrophobicity of specific regions in Im9, altering its sequence to incorporate three residues from Im7, or changing pH resulted in the detectable accumulation of an Im9 intermediate[29; 30]. A similar phenomenon was observed for members of the homeodomain family. Three members of the family appeared two-state while the engrailed homeodomain (En-HD) folded through an

intermediate[31]. When simulations suggested that c-Myb, another member of the homeodomain family, also populate an intermediate, the authors increased the helix propensity of helix II by removing a single proline, and three-state folding emerged[32]. A switch in apparent mechanism by changing a few residues or conditions has been observed for a few other proteins[33; 34; 35], suggesting an underlying roughness to the energy landscape of proteins that appear to be two-state. This property is also implied even in a case in which two members of a protein family both fold through an observable intermediate. Two members of the intracellular lipid-binding protein family fold through distinct intermediates differing in their secondary structure content[36]. Swapping a core residue in these proteins results in a switch in the type of intermediate that each protein populates. These studies point to the pliable and rugged nature of the energy landscape where the accumulation or structural properties of intermediates can be manipulated by simple changes.

The development of new techniques with increased sensitivity and time resolution has also brought to light intermediates that otherwise might elude detection. Ultra-rapid mixing techniques uncovered an early folding intermediate in the small (86 amino acids) four-helix protein acyl-CoA binding protein, previously thought to be two-state[37]. Roder and co-workers utilized a continuous flow mixer with a dead time of 70 μs combined with detection by Forster resonance energy transfer (FRET). Another exciting development has been in the field of nuclear magnetic resonance, where relaxation dispersion methods are now capable of detecting non-native species present down to 0.5%[38]. Residue-specific relaxation dispersion profiles depend on the population of two interconverting species, the interconversion rates, and the chemical shift difference between the two states. Thus, they can provide invaluable structural, thermodynamic, and kinetic information. These techniques have been used to uncover intermediates for ACBP[18] and Fyn-SH3[39].

Intermediates as misfolded species

When the presence of an intermediate is established, what role does this partially folded species play in folding? This appears to be protein-dependent, and evidence for harmful and helpful intermediates exists. The Im7 intermediate discussed above is actually a misfolded species. Eight destabilizing mutations at the interface of helices I and II, which are structured in the intermediate, result in an increased rate of folding to the native state from the intermediate, indicating that substantial non-native interactions are formed early in this region of the protein[40]. Interestingly, however, this species was shown to be on-pathway[41].

The 159-residue major birch pollen antigen Bet v 1, a water channel, has also been observed to populate a misfolded intermediate[33]. This intermediate, however, is off-pathway; the protein must go back to the unfolded state before it folds to the native state. A single point mutant, Y120W, which does not change the stability of the protein, eliminates the misfolded intermediate. This mutant was designed to eliminate non-native interactions present in the partially folded form. The difference in the folding rates of these two proteins is about an order of magnitude: 13 s^{-1} for the wild type protein and 125 s^{-1} in the Y120W variant. The authors speculate that in a biological context, this magnitude of increase in the folding rate has no serious implications, which is why there has been no selection against this residue. Non-native packing has also been reported to stabilize the intermediates of α-lactalbumin[42], SH3 domains[43], lysozyme[44], and CD2.d1[45].

NMR structures have been determined for a mutant of the engrailed homeodomain that populates its kinetic intermediate at equilibrium and for an engineered mimic of a postulated unfolding intermediate of Rd-apocytochrome b_{562}, and both are marked by the presence of non-native interactions. Misfolded intermediates appear to be a common feature and do not necessarily represent a hindrance to efficient folding. They very well may speed folding if they minimize the search for the native state enough to overcome the cost of breaking the non-native interactions.

Intermediates as beneficial species

There are also examples of intermediates that are part of a hierarchical folding mechanism in which interactions increase in number and strength as folding progresses. The kinetic intermediates of *E. coli* RNase H[35] and of Trp cage[34], a 20-residue "miniprotein," can be eliminated by single point mutations. The two-state proteins fold more slowly, indicating that population of the intermediate promotes efficient folding. Interactions that stabilize the intermediate must also stabilize the transition state to a greater extent to result in a rate enhancement. Phi-values also increase from the intermediate to the transition state for *E. coli* RNase H[35; 46], and the same is true for barnase[47; 48]. A hierarchical mechanism has also been invoked for the four-helix FF domain from HYPA/FBP11[49]. Intermediates that are populated at equilibrium and are believed to correspond to kinetic intermediates, such as the molten globule forms of *E. coli* RNase H[10; 50], apomyoglobin[8; 51], and cytochrome c[9] are also sited as examples of hierarchical folding. These intermediates are characterized by native-like secondary structure formation but lack fixed tertiary contacts. However, there is evidence that the kinetic intermediate of apomyoglobin is stabilized by non-native interactions in addition to native contacts[52]. Hydrogen exchange pulse-labeling experiments were performed on several mutants, and the results point to the formation of a burst phase intermediate with an overall native-like topology but with translocation of the H helix by one helical turn towards its N terminus to maximize hydrophobic interactions with helix G. These non-native interactions must be resolved before the protein can proceed to the native state.

Ruggedness of the energy landscape

The examples discussed above suggest an underlying complexity and roughness to the energy landscapes of even small proteins whose folding appears two-state. This may be because evolution acts on the energy landscape under functional constraints. Many of the residues that form non-native contacts in the Im7 intermediate, for instance, are also crucial for recognizing and inactivating colicin toxins[53]. Residues crucial to function also appear to slow folding for the Pin1 WW domain[54], the β-trefoil motifs interleukin-1β and interleukin-1Ra[55], and Fyn SH3[56]. Fyn SH3 is particularly interesting because folding can proceed fastest through a non-native conformation. A series of mutations was made at the highly conserved Gly 48, all of which destabilize the protein yet speed folding. There is a strong correlation between rates and beta sheet propensity, even though Gly48 is in a non-beta strand backbone conformation in the native state. However, the mutants all demonstrated a reduced affinity for the target peptide. Folding efficiency may therefore be sacrificed in light of functional requirements, resulting in increased ruggedness on the energy landscape.

1.4 FOLDING STUDIES OF *E. coli* RNase H

The majority of work presented in this thesis explores the presence and role of the kinetic intermediate of ribonuclease HI (RNase H) from *Escherichia coli*. *E. coli* RNase H is an ideal system for examining the role of intermediates on protein energy landscapes. As a small (MW = 17.5 kDa), monomeric enzyme with an $\alpha + \beta$ fold and no disulphide bonds[57], it is well-suited for folding studies. RNase H cleaves the RNA strand of RNA:DNA hybrids. Studies from the Marqusee laboratory have focused on the cysteine-free variant of the protein, in which three cysteines have been replaced by alanine. These mutations do not affect the structure or activity of the enzyme[58]. The crystal structure of the cysteine-free variant is shown in Figure 1.3 and is henceforth referred to as wild type *E. coli* RNase H.

Early work on the folding pathway of RNase H indicated that the protein folds through a kinetic intermediate[10]. The presence of a transiently populated, partially folded intermediate was supported by several observations. First, when folding was monitored by stopped-flow circular dichroism, ~80 % of the total signal change between unfolded and folded protein occurred within the mixing time of the instrument (12 ms). This burst phase was followed by a much slower, observable signal decay that can be described by a single exponential. Furthermore, the burst phase amplitude changed cooperatively with denaturant concentration. Finally, the chevron for RNase H displays a rollover in the folding limb, which is consistent with the accumulation of a kinetic intermediate.

The structure of the kinetic intermediate was probed by pulse-labeling hydrogen exchange[10]. At the earliest time point examined, 14 ms, amide protons in the core of the molecule, namely helices A and D and strand 4, are well protected and therefore presumed to be structured at this point during refolding. By stopped-flow circular dichroism, the intermediate is highly populated at 14 ms; thus these secondary structure elements are well formed in this species. Probes in the rest of the beta strands, the periphery of the molecule, did not show protection at this early time point and gained protection at a rate of 1 s^{-1}, which is consistent with the slower phase of folding observed by stopped-flow CD. Thus it is presumed that these regions of the protein are not well formed in the kinetic intermediate. The structural properties of the intermediate are mirrored in the partially folded forms of RNase H found to exist at equilibrium. Hydrogen exchange protection patterns of the acid-state molten globule[59] and a partially folded form detected under native conditions[60] are very similar. Order of structure formation appears to be thermodynamically driven in a hierarchical manner in which the most stable regions in the core of the protein fold first.

The hierarchical mechanism is also supported by ϕ-value analysis[46]. A series of four mutations were examined in RNase H: two in the core of the protein and two in the periphery. The reaction coordinate for wild type *E. coli* RNase H and these four mutants are depicted in Figure 1.4. The mutations in the core, I53A and Q105G, slowed folding while leaving the unfolding rate unaffected. This results in a high ϕ-value, indicating that these residues are well-formed in the transition state. The ϕ-values for the mutations in the periphery were slightly harder to interpret, but the qualitative shifts in the chevron were clear: the stabilizing effect of the D10A mutation, which relieves charge repulsion, and the destabilizing effect of the R27A mutation were manifested primarily in the unfolding limb, with slight shifts to the folding limbs. Thus, the periphery region is relatively unstructured in the rate-limiting transition state. ϕ-values for the intermediate for all four mutations were also calculated and were observed to increase from the

intermediate to the transitions state. Interactions formed in the intermediate persist and strengthen in transition state. In Chapter 2 of this thesis, I describe my work using ϕ-value analysis to examine the relative contributions of topology and regional stability to the folding mechanism. I found that topology actually dictates the order of structure formation while the core stability determines how fast the protein can fold.

These experiments provide a general view of the progression of the molecule as it folds: *E. coli* RNase H folds through an intermediate composed of structure in the core of the molecule and a relatively unstructured periphery, which gains structure only after the rate-limiting step. Many questions remained as to the nature of the interactions stabilizing the folding core. The tertiary interactions may be loose and fluctuating, or they may be fixed and native-like. To investigate this, a series of mutations at residue 53 in Helix A were made, and their affect on the kinetic intermediate, transition state, and native state were compared[35]. Isoleucine 53, shown in red in Figure 1.3, was replaced with four hydrophobic residues (Leu, Phe, Ala, Val) and one polar residue (Asp). The mutations affect the intermediate to varying degrees. Both I53V and I53L, relatively conservative mutations, fold through a kinetic intermediate with a stability very similar to that of the wild type intermediate. I53A, on the other hand, a more drastic removal of the hydrophobic side chain, destabilizes the intermediate to the same extent it destabilizes the native state. This reveals that hydrophobic interactions are largely responsible for stabilizing the intermediate. Also, as long as hydrophobicity is somewhat conserved, the exact identity of the amino acid has little effect, suggesting that side-chain packing is not fixed in the intermediate. This cannot be solely responsible for the stability of the intermediate, however: I53F, a mutation that conserves hydrophobicity but alters steric packing interactions, destabilizes the intermediate by a significant amount. While the subtle change in sterics (I53L) is completely tolerated in the intermediate, the more drastic introduction of a large aromatic residue (I53F) is not. While the stability of the intermediate largely arises from non-specific hydrophobic interactions, packing interactions must play a role to some extent.

The largest affect on the intermediate is seen with the I53D mutant: this protein folds in an apparent two-state manner, without the detectable presence of an intermediate. This mutation destabilizes the protein by 4.2 kcal/mol. A destabilization of the intermediate of more than 3.5 kcal/mol would result in a species higher in energy than the unfolded state, and we would not expect to be able to see the accumulation of the intermediate. Adding sodium sulfate results in a return of a detectable intermediate for I53D, suggesting this is indeed the case. However, it is also possible that the two-state I53D folds through a completely different trajectory (rate-limiting transition state) than the wild-type protein.

The affect of these mutations on the transition state was also examined. I53L has a ϕ-value close to zero; this conservative mutation does not change the energetics of the transition state. I53F had a greater ϕ-value for the transition state than the ϕ-value for the intermediate, indicating a strengthening of interactions from the intermediate to the transition state, where specific side-chain interactions apparently become more important. I53A had similarly destabilizing affect on both, again highlighting the importance of hydrophobic interactions. The same interactions that stabilize the intermediate stabilize the transition state to a greater extent, resulting in an acceleration of folding. Populating an intermediate is therefore productive for efficient folding. Chapters 3 and 4 describe

my work to create an equilibrium mimic of the kinetic intermediate in order to study this important species directly.

Single-molecule experiments conclusively demonstrated that RNase H folds through an on-pathway, obligatory intermediate[61]. In these experiments, single protein molecules were unfolded by force in the optical tweezers. To carry out these experiments, cysteines were introduced at the N- and C-termini of the protein, and DNA "handles" with thiols at one end were covalently attached to the protein. The handles were further functionalized on the opposite ends with antigens, and these in turn interact with antibody-coated beads. The micron size beads can be manipulated with a micropipette and a laser trap. RNase H was observed to unfold in a two-state manner and refold through an intermediate. When sitting at a single force and observing the extension change of the molecule, the protein was observed to "hop" between the unfolded and intermediate states. Occasionally, the molecule would proceed to the native state directly from the intermediate, and hopping was halted, demonstrating that the intermediate is on-pathway. No such intermediate was observed for I53D RNase H, suggesting that the intermediate observed in these single-molecule forced unfolding experiments correlates to the species observed in all the previous bulk experiments.

1.5 COMPARISON OF THE ENERGY LANDSCAPES OF *E. coli* AND *T. thermophilus* RNases H

Comparing the folding trajectories of closely related proteins can offer considerable insight into the factors that determine if an intermediate is populated and what barriers are traversed on the energy landscape. It is further enlightening to compare the folding of the protein from a thermophilic organism to that of its homologue from a mesophile. In addition to probing the role of topology in determining the folding trajectory, such pairwise comparisons allow us to ask how proteins with similar sequences and virtually identical structures exhibit vastly different thermodynamic properties.

To this end, the energy landscape of RNase H from *Thermus thermophilus* has been extensively studied and compared to that of RNase H from *E. coli*. *T. thermophilus* grows at 68.5 ^{o}C, while the optimal growth temperature for *E. coli* is 37 ^{o}C. The midpoint of thermal denaturation of RNase H from *T. thermophilus* is 86 ^{o}C, while that of the *E. coli* protein is 66 ^{o}C[62]. Therefore, the protein from the thermophile must be folded and must function at temperatures where the protein from the mesophile is mostly unfolded. The two proteins share 52 % sequence identity, and their crystal structures are nearly indistinguishable, with an rms deviation between alpha carbons in secondary structural regions of 0.95 Å[63]. Given their identical architectures, it is not surprising that many general similarities arise. Native state hydrogen exchange experiments revealed that although all residues exchange at a higher free energy in the thermophilic protein, the general distribution of stability in *T. thermophilus* RNase H is similar to that of *E. coli* RNase H[64]. Each partially folded form that was identified has a homologous counterpart in *E. coli*, with the exception of the local unfolding of the E helix at a lower free energy than the beta strands in the periphery. In *E. coli* RNase H, the E helix clusters with the beta sheets. These results suggest that a finely tuned balance of interactions has been preserved. Furthermore, *T. thermophilus* RNase H also populates a burst phase kinetic intermediate that is similar in structure to that populated by *E. coli* RNase H[65]. Quenched

flow hydrogen exchange experiments show protection at the earliest time points in the core of the molecule, with the periphery gaining protection on the order of the observable phase. The two proteins fold and unfold with similar overall rates. Unlike *E. coli* RNase H, however, the folding of *T. thermophilus* RNase H appears limited by proline isomerization.

What, then, are the important differences that confer such different thermodynamic behavior? Changing three residues in *E. coli* RNase H to their corresponding identity in the *T. thermophilus* protein increased T_m[66; 67; 68]. The effect of each mutation was additive, and the three changes together did not entirely account for the difference in T_m. NMR studies have detected some differences in the backbone dynamics of the two proteins[69; 70]. The most crucial difference between the two proteins was revealed in a comparison of their stability curves. *T. thermophilus* RNase H has a much lower change in heat capacity upon unfolding, ΔC_p[62]. The lower ΔC_p directly contributes to the increase in T_m for the thermophilic protein. This property was found to track with the core of the protein in a study in which chimeras of the proteins were created[71]; the folding core must play a key role in determining the thermodynamic properties of each protein. Differential scanning calorimetry was used to verify that the low ΔC_p arises from residual structure in the unfolded state of *T. thermophilus* RNase H, and the hydrophobic nature of the residual structure was demonstrated by mutagenesis[72].

Structural differences in the unfolded states of these two proteins appear to be important for determining their thermodynamic behavior. Given that these differences originate in the core, the structured region of the intermediate, it is possible that there may be interesting differences in the properties of the intermediate that also contribute. The work presented in Chapter 4 explores this possibility.

1.6 REPEAT PROTEINS AS MODEL SYSTEMS

The studies discussed to this point have been carried out on globular proteins, rich in sequence-distance contacts. In such systems, it is not hard to imagine how changes at a particular site in the structure can be propagated to the rest of the protein. This provides somewhat of a structural basis for the cooperativity that is observed for equilibrium unfolding transitions and which largely obscures our ability to study partially folded species. Repeat proteins, on the other hand, are tandem arrays of repeating structural units and have no long-range contacts[73]. The elongated structures of various repeat proteins are shown in Figure 1.5. In contrast to globular domains, the majority of native contacts could be maintained in the face of partial unfolding. In spite of this modularity, many exhibit highly cooperative equilibrium unfolding transitions that can be described by the two-state approximation. Repeat proteins therefore offer a unique architecture for exploring the extent of cooperativity and roughness on the energy landscape.

Ankyrin repeat proteins

Approximately 30% of proteins encoded in eukaryotic genomes contain repeated sequences[73]. The repeats can be composed of all β, all α, or mixed α and β secondary structure. Ankyrin repeat proteins make up one class of repeat proteins whose folding has been extensively studied[74]. The ankyrin repeat consists of two antiparallel helices connected by a beta turn[75]. Ankyrin repeat proteins participate in a diverse set of biological roles; they largely mediate protein-protein interactions in cellular processes such as transcriptional regulation, cell cycle regulation, and signaling[76; 77].

Thermodynamic studies of ankyrin repeat proteins reveal cooperative unfolding transitions for proteins containing up to six repeats[77; 78; 79]. A larger 12-ankyrin repeat fragment of ankyrin-R is known to populate intermediates at equilibrium[80].

Although the equilibrium transitions of smaller ankyrin repeat proteins are often highly cooperative, many have been observed to populate kinetic intermediates. $p16^{INK4a}$, which consists of four ankyrin repeats, and $p19^{INK4d}$, containing five ankyrin repeats, fold through on-pathway kinetic intermediates[81; 82], as does the Notch ankyrin domain, a seven-ankyrin repeat protein[83]. Though it might be expected that these proteins reach the native state via multiple pathways, ϕ value analysis suggests a discrete folding pathway for $p16^{INK4a}$ [84] and the Notch ankyrin domain[85]. Structure formation proceeds through a two or three-repeat portion of these proteins. The case of myotrophin, another four-repeat protein, is less clear. Two explanations have been invoked to interpret the kinetic data; one describes a two-state transition with a broad energy barrier[79], and one proposes parallel pathways and a rerouting of folding when the C-terminal repeats are destabilized[86].

The origin and extent of energetic coupling in repeat proteins can be more rigorously examined in single-molecule force spectroscopy folding studies, for which they are ideally suited. Force can be applied along the axis on which the protein is modular, and it seems plausible that repeats will unfold individually, accessing multiple intermediates along the way. Furthermore, it has been suggested that ankyrin repeat proteins are a component of a gated spring in the hair cells of vertebrates, making their response to mechanical force directly relevant for function[87; 88; 89].

Three ankyrin repeat proteins have been studied in the atomic force microscope (AFM). N6C is an array of six consensus ankyrin repeats flanked by solubilizing "caps." It exhibits stepwise unfolding of single repeat units in the AFM[90]. This stepwise behavior was also observed for ankyrin-B, a large, natural protein containing 24 ankyrin repeats[91]. Gankyrin, another naturally-occurring protein composed of seven repeats, demonstrated slightly different behavior[92]. It was observed to unfold largely in units of half or one full repeat, but unfolding transitions containing two repeats or more were also detected. It appears that the pathways accessed under mechanical force are more heterogeneous and less cooperative as compared to bulk studies. These differences may arise from the use of force as a denaturant as opposed to chemicals that act globally on the protein chain. The ensemble averaging inherent to bulk experiments may also mask some of the complexity of repeat protein folding. There are many unanswered questions as to the malleability of the energy landscapes of repeat proteins and what relevance it has to their function.

1.7 SUMMARY OF WORK IN THIS THESIS

Chapter 2

This chapter discusses the relative contributions of regional stability and topology in determining the folding pathway of *E. coli* RNase H using an idea originally put forth by Erik Miller[93]. I used phi-value analysis to compare the structures of the rate-limiting transition states of the wild-type protein, which folds in a three-state mechanism, and the variant I53D RNase H, which folds in an apparent two-state mechanism. Having severely destabilized the region of the molecule through which folding proceeds with this single residue change in the core, is protein re-routed to incorporate elements of the

periphery? Two mutations in the core, W85A and Q105G, and two mutations in the periphery, F8A and I25A, were examined in the three-state and two-state backgrounds. We found that altering local stability does not result in a change in the order of structure formation in the molecule; the core residues are highly structured in the rate-limiting transition states of both proteins, and the periphery residues contribute to an equally small degree. Topology must play a key role in determining the folding pathway of RNase H, and the formation of a species in which the core is folded while the periphery is not is critical for folding. This chapter appears in the Journal of Molecular Biology with Chapter 3 as an accompanying manuscript.

Chapter 3

In light of the general importance of intermediate formation in the efficient folding of *E. coli* RNase H, we set out to study this species more directly. In work begun by Geoff Horner, I set out to create an equilibrium population of the kinetic intermediate by selectively destabilizing the native state[94]. The native state can, in principle, be selectively destabilized by introducing mutations in the periphery of the protein, the region that is largely unfolded in the intermediate. Surprisingly, the introduction of a single mutation into the periphery, I25A, was sufficient to accomplish this goal. Examining the effect of the I25A mutation in a more stable background, D10A RNase H, allowed us to map the populations of the native, intermediate, and unfolded states of the I25A variant as a function of denaturant. The folded regions of the intermediate were undetectable in a series of HSQC's, suggesting that this species is dynamic on the timescale of NMR detection. This property sets the *E. coli* RNase H intermediate apart from several models in the literature and may be important for its role in promoting efficient folding. This chapter appears in the Journal of Molecular Biology in accompaniment with Chapter 2.

Chapter 4

To more fully populate the intermediate of *E. coli* RNase H at equilibrium, I continued to introduce destabilizing mutations into the periphery. The more complete population of the intermediate allows for a more rigorous examination of the dynamic properties of this species. Additionally, to compare this species to its counterpart in *T. thermophilus* RNase H, I took the same protein engineering approach with the protein from the thermophilic organism. Recent NMR studies of a fragment of *T. thermophilus* RNase H corresponding to the folding core demonstrate that it is a well-packed and rigid subset of the native state. The work in this chapter seeks to understand the origin of the apparent difference in dynamic behavior between the intermediates of the two homologous proteins. CD and NMR experiments indicate that we have successfully generated mimics of the kinetic intermediates of these two proteins. Comparison of the HSQC's of the various constructs provides interesting questions to pursue. The results presented here are preliminary, and research is ongoing.

Chapter 5

This chapter describes the use of the optical tweezers, a single-molecule mechanical technique, to study the folding of the Notch ankyrin domain, a repeat protein. Repeat proteins, with their modular structures that lack long-range contacts, are ideal systems for examining the origin and extent of cooperativity in protein folding. In spite of their modular architectures, many demonstrate cooperative equilibrium unfolding transitions. I found that the Notch ankyrin domain exhibits increased heterogeneity in forced

unfolding pathways compared to bulk studies, but the limit of this decoupling is also evident. Notably, in contrast to the reported behavior of ankyrin repeat proteins in the atomic force microscope, stepwise unfolding of repeats is not observed. Additionally, the smallest unit stable to force appears to be three repeats, demonstrating that individual repeats maintain a level of cooperativity. A comparison of these results to the unfolding behavior of the Notch ankyrin domain in the AFM provides insight into the apparent difference in solution and forced unfolding of ankyrin repeats. Since these data were collected, there have been significant improvements in optical tweezer instrumentation, including greatly enhanced time resolution. The experiments described in this chapter are currently being carried out on the new optical tweezer set-up.

1.8 REFERENCES

1. Anfinsen, C. B. & Haber, E. (1961). Studies on the reduction and re-formation of protein disulfide bonds. *J Biol Chem* **236**, 1361-3.
2. Levinthal, C. (1968). Are there pathways for protein folding? *Journal of Chimica and Physics* **65**, 44-45.
3. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-9.
4. Chan, H. S. & Dill, K. A. (1998). Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins* **30**, 2-33.
5. Uversky, V. N. The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *J Biomed Biotechnol* **2010**, 568068.
6. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold Des* **3**, R81-91.
7. Clarke, J., Itzhaki, L. S. & Fersht, A. R. (1997). Hydrogen exchange at equilibrium: a short cut for analysing protein-folding pathways? *Trends Biochem Sci* **22**, 284-7.
8. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892-6.
9. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192-7.
10. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* **4**, 298-304.
11. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* **29**, 213-38.
12. Fersht, A. R. (1999). *Structure and Mechanism in Protein Science*, W. H. Freeman and Company.
13. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-35.
14. Baldwin, R. L. (1995). On-pathway versus off-pathway folding intermediates. *Fold Des* **1**, R1-8.
15. Zhou, Z., Huang, Y. & Bai, Y. (2005). An on-pathway hidden intermediate and the early rate-limiting transition state of Rd-apocytochrome b562 characterized by protein engineering. *J Mol Biol* **352**, 757-64.
16. Kato, H., Vu, N. D., Feng, H., Zhou, Z. & Bai, Y. (2007). The folding pathway of T4 lysozyme: an on-pathway hidden folding intermediate. *J Mol Biol* **365**, 881-91.
17. Feng, H., Vu, N. D. & Bai, Y. (2005). Detection of a hidden folding intermediate of the third domain of PDZ. *J Mol Biol* **346**, 345-53.
18. Teilum, K., Poulsen, F. M. & Akke, M. (2006). The inverted chevron plot measured by NMR relaxation reveals a native-like unfolding intermediate in acyl-CoA binding protein. *Proc Natl Acad Sci U S A* **103**, 6877-82.

19. Sanchez, I. E. & Kiefhaber, T. (2003). Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys Chem* **100**, 397-407.
20. Otzen, D. E., Kristensen, O., Proctor, M. & Oliveberg, M. (1999). Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry* **38**, 6499-511.
21. Sanchez, I. E. & Kiefhaber, T. (2003). Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J Mol Biol* **327**, 867-84.
22. Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122-6.
23. Zhou, Z., Feng, H., Ghirlando, R. & Bai, Y. (2008). The high-resolution NMR structure of the early folding intermediate of the Thermus thermophilus ribonuclease H. *J Mol Biol* **384**, 531-9.
24. Feng, H., Zhou, Z. & Bai, Y. (2005). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc Natl Acad Sci U S A* **102**, 5026-31.
25. Kato, H., Feng, H. & Bai, Y. (2007). The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J Mol Biol* **365**, 870-80.
26. Gsponer, J., Hopearuoho, H., Whittaker, S. B., Spence, G. R., Moore, G. R., Paci, E., Radford, S. E. & Vendruscolo, M. (2006). Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc Natl Acad Sci U S A* **103**, 99-104.
27. Whittaker, S. B., Spence, G. R., Gunter Grossmann, J., Radford, S. E. & Moore, G. R. (2007). NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J Mol Biol* **366**, 1001-15.
28. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **286**, 1597-608.
29. Gorski, S. A., Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2001). Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. *J Mol Biol* **312**, 849-63.
30. Friel, C. T., Beddard, G. S. & Radford, S. E. (2004). Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J Mol Biol* **342**, 261-73.
31. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W., DeMarco, M. L., Daggett, V. & Fersht, A. R. (2003). Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A* **100**, 13286-91.
32. White, G. W., Gianni, S., Grossmann, J. G., Jemth, P., Fersht, A. R. & Daggett, V. (2005). Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J Mol Biol* **350**, 757-75.

33. Mogensen, J. E., Ipsen, H., Holm, J. & Otzen, D. E. (2004). Elimination of a misfolded folding intermediate by a single point mutation. *Biochemistry* **43**, 3357-67.
34. Neuweiler, H., Doose, S. & Sauer, M. (2005). A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate. *Proc Natl Acad Sci U S A* **102**, 16650-5.
35. Spudich, G. M., Miller, E. J. & Marqusee, S. (2004). Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol* **335**, 609-18.
36. Dalessio, P. M., Boyer, J. A., McGettigan, J. L. & Ropson, I. J. (2005). Swapping core residues in homologous proteins swaps folding mechanism. *Biochemistry* **44**, 3082-90.
37. Teilum, K., Maki, K., Kragelund, B. B., Poulsen, F. M. & Roder, H. (2002). Early kinetic intermediate in the folding of acyl-CoA binding protein detected by fluorescence labeling and ultrarapid mixing. *Proc Natl Acad Sci U S A* **99**, 9807-12.
38. Neudecker, P., Lundstrom, P. & Kay, L. E. (2009). Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys J* **96**, 2045-54.
39. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586-90.
40. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* **9**, 209-16.
41. Capaldi, A. P., Shastry, M. C., Kleanthous, C., Roder, H. & Radford, S. E. (2001). Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol* **8**, 68-72.
42. Mok, K. H., Nagashima, T., Day, I. J., Hore, P. J. & Dobson, C. M. (2005). Multiple subsets of side-chain packing in partially folded states of alpha-lactalbumins. *Proc Natl Acad Sci U S A* **102**, 8899-904.
43. Neudecker, P., Zarrine-Afsar, A., Choy, W. Y., Muhandiram, D. R., Davidson, A. R. & Kay, L. E. (2006). Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J Mol Biol* **363**, 958-76.
44. Radford, S. E., Dobson, C. M. & Evans, P. A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302-7.
45. Mason, J. M., Cliff, M. J., Sessions, R. B. & Clarke, A. R. (2005). Low energy pathways and non-native interactions: the influence of artificial disulfide bridges on the mechanism of folding. *J Biol Chem* **280**, 40494-9.
46. Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* **6**, 825-31.
47. Serrano, L., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol* **224**, 805-18.

48. Matouschek, A., Serrano, L. & Fersht, A. R. (1992). The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol* **224**, 819-35.
49. Korzhnev, D. M., Religa, T. L., Lundstrom, P., Fersht, A. R. & Kay, L. E. (2007). The folding pathway of an FF domain: characterization of an on-pathway intermediate state under folding conditions by (15)N, (13)C(alpha) and (13)C-methyl relaxation dispersion and (1)H/(2)H-exchange NMR spectroscopy. *J Mol Biol* **372**, 497-512.
50. Chamberlain, A. K. & Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry* **37**, 1736-42.
51. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**, 1544-8.
52. Nishimura, C., Dyson, H. J. & Wright, P. E. (2006). Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J Mol Biol* **355**, 139-56.
53. Friel, C. T., Smith, D. A., Vendruscolo, M., Gsponer, J. & Radford, S. E. (2009). The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat Struct Mol Biol* **16**, 318-24.
54. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M. & Kelly, J. W. (2006). Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci U S A* **103**, 10648-53.
55. Gosavi, S., Whitford, P. C., Jennings, P. A. & Onuchic, J. N. (2008). Extracting function from a beta-trefoil folding motif. *Proc Natl Acad Sci U S A* **105**, 10384-9.
56. Di Nardo, A. A., Korzhnev, D. M., Stogios, P. J., Zarrine-Afsar, A., Kay, L. E. & Davidson, A. R. (2004). Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation. *Proc Natl Acad Sci U S A* **101**, 7954-9.
57. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from E. coli. *Nature* **347**, 306-9.
58. 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.
59. Dabora, J. M., Pelton, J. G. & Marqusee, S. (1996). Structure of the acid state of Escherichia coli ribonuclease HI. *Biochemistry* **35**, 11951-8.
60. Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* **3**, 782-7.
61. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057-60.
62. Hollien, J. & Marqusee, S. (1999). A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* **38**, 3831-6.

63. Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H. & Morikawa, K. (1993). Crystal structure of ribonuclease H from Thermus thermophilus HB8 refined at 2.8 A resolution. *J Mol Biol* **230**, 529-42.
64. Hollien, J. & Marqusee, S. (1999). Structural distribution of stability in a thermophilic enzyme. *Proc Natl Acad Sci U S A* **96**, 13674-8.
65. Hollien, J. & Marqusee, S. (2002). Comparison of the folding processes of T. thermophilus and E. coli ribonucleases H. *J Mol Biol* **316**, 327-40.
66. 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**, 6171-8.
67. Kimura, S., Kanaya, S. & Nakamura, H. (1992). Thermostabilization of Escherichia coli ribonuclease HI by replacing left-handed helical Lys95 with Gly or Asn. *J Biol Chem* **267**, 22014-7.
68. Kimura, S., Nakamura, H., Hashimoto, T., Oobatake, M. & Kanaya, S. (1992). Stabilization of Escherichia coli ribonuclease HI by strategic replacement of amino acid residues with those from the thermophilic counterpart. *J Biol Chem* **267**, 21535-42.
69. Butterwick, J. A. & Palmer, A. G., 3rd. (2006). An inserted Gly residue fine tunes dynamics between mesophilic and thermophilic ribonucleases H. *Protein Sci* **15**, 2697-707.
70. Butterwick, J. A., Patrick Loria, J., Astrof, N. S., Kroenke, C. D., Cole, R., Rance, M. & Palmer, A. G., 3rd. (2004). Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. *J Mol Biol* **339**, 855-71.
71. Robic, S., Berger, J. M. & Marqusee, S. (2002). Contributions of folding cores to the thermostabilities of two ribonucleases H. *Protein Sci* **11**, 381-9.
72. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M. & Marqusee, S. (2003). Role of residual structure in the unfolded state of a thermophilic protein. *Proc Natl Acad Sci U S A* **100**, 11345-9.
73. Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. (1999). A census of protein repeats. *J Mol Biol* **293**, 151-60.
74. Barrick, D., Ferreiro, D. U. & Komives, E. A. (2008). Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Curr Opin Struct Biol* **18**, 27-34.
75. Sedgwick, S. G. & Smerdon, S. J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem Sci* **24**, 311-6.
76. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* **13**, 1435-48.
77. Li, J., Mahajan, A. & Tsai, M. D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* **45**, 15168-78.
78. Zweifel, M. E. & Barrick, D. (2001). Studies of the ankyrin repeats of the Drosophila melanogaster Notch receptor. 2. Solution stability and cooperativity of unfolding. *Biochemistry* **40**, 14357-67.
79. Lowe, A. R. & Itzhaki, L. S. (2007). Biophysical characterisation of the small ankyrin repeat protein myotrophin. *J Mol Biol* **365**, 1245-55.

80. Werbeck, N. D. & Itzhaki, L. S. (2007). Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. *Proc Natl Acad Sci U S A* **104**, 7863-8.
81. Tang, K. S., Guralnick, B. J., Wang, W. K., Fersht, A. R. & Itzhaki, L. S. (1999). Stability and folding of the tumour suppressor protein p16. *J Mol Biol* **285**, 1869-86.
82. Zeeb, M., Rosner, H., Zeslawski, W., Canet, D., Holak, T. A. & Balbach, J. (2002). Protein folding and stability of human CDK inhibitor p19(INK4d). *J Mol Biol* **315**, 447-57.
83. Mello, C. C., Bradley, C. M., Tripp, K. W. & Barrick, D. (2005). Experimental characterization of the folding kinetics of the notch ankyrin domain. *J Mol Biol* **352**, 266-81.
84. Tang, K. S., Fersht, A. R. & Itzhaki, L. S. (2003). Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure* **11**, 67-73.
85. Bradley, C. M. & Barrick, D. (2006). The notch ankyrin domain folds via a discrete, centralized pathway. *Structure* **14**, 1303-12.
86. Lowe, A. R. & Itzhaki, L. S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc Natl Acad Sci U S A* **104**, 2679-84.
87. Gao, M., Sotomayor, M., Villa, E., Lee, E. H. & Schulten, K. (2006). Molecular mechanisms of cellular mechanics. *Phys Chem Chem Phys* **8**, 3692-706.
88. Howard, J. & Bechstedt, S. (2004). Hypothesis: a helix of ankyrin repeats of the NOMPC-TRP ion channel is the gating spring of mechanoreceptors. *Curr Biol* **14**, R224-6.
89. Walker, R. G., Willingham, A. T. & Zuker, C. S. (2000). A Drosophila mechanosensory transduction channel. *Science* **287**, 2229-34.
90. Li, L., Wetzel, S., Pluckthun, A. & Fernandez, J. M. (2006). Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. *Biophys J* **90**, L30-2.
91. Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V. & Marszalek, P. E. (2006). Nanospring behaviour of ankyrin repeats. *Nature* **440**, 246-9.
92. Serquera, D., Lee, W., Settanni, G., Marszalek, P. E., Paci, E. & Itzhaki, L. S. Mechanical unfolding of an ankyrin repeat protein. *Biophys J* **98**, 1294-301.
93. Connell, K. B., Miller, E. J. & Marqusee, S. (2009). The folding trajectory of RNase H is dominated by its topology and not local stability: a protein engineering study of variants that fold via two-state and three-state mechanisms. *J Mol Biol* **391**, 450-60.
94. Connell, K. B., Horner, G. A. & Marqusee, S. (2009). A single mutation at residue 25 populates the folding intermediate of E. coli RNase H and reveals a highly dynamic partially folded ensemble. *J Mol Biol* **391**, 461-70.
95. Kloss, E., Courtemanche, N. & Barrick, D. (2008). Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* **469**, 83-99.

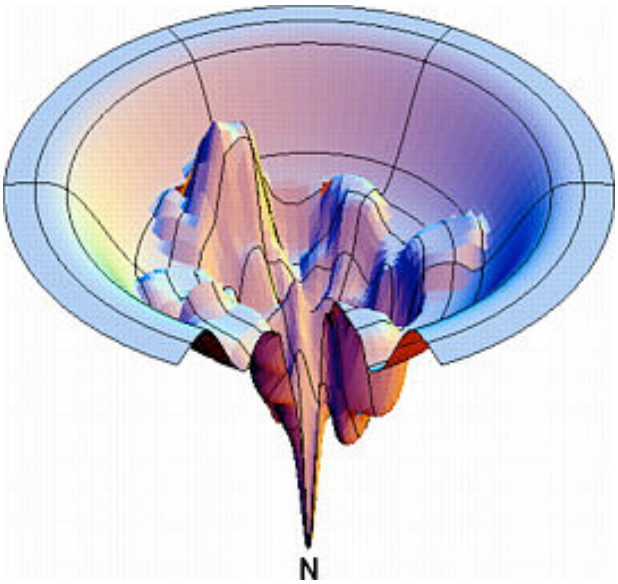
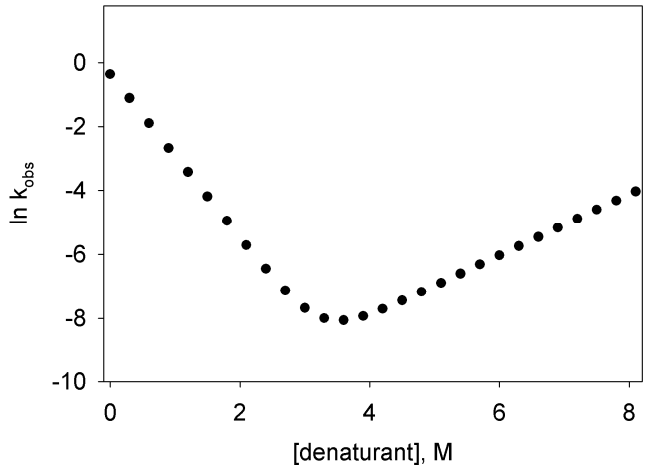


Figure 1.1: A schematic representation of the energy landscape of a hypothetical protein. In this visualization, the horizontal axes represents the various conformations of the protein, and the vertical axis represents the internal free energy of the protein chain. As the protein folds, it moves from the upper plain of the landscape, representing the various unfolded conformations, to the global energy minimum of the native state. This hypothetical landscape was generated by the Dill lab[3].

(a)



(b)

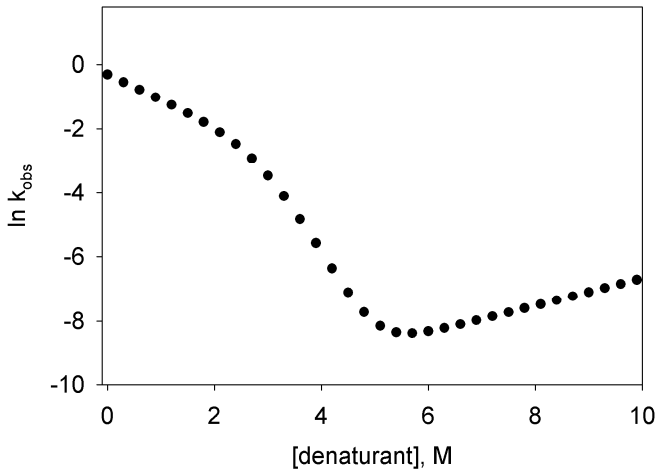


Figure 1.2: Simulated chevron plots. (a) The classic V-shaped chevron plot of a two-state folding protein. (b) The chevron plot of a protein that folds through a burst phase intermediate, displaying rollover in the folding limb.

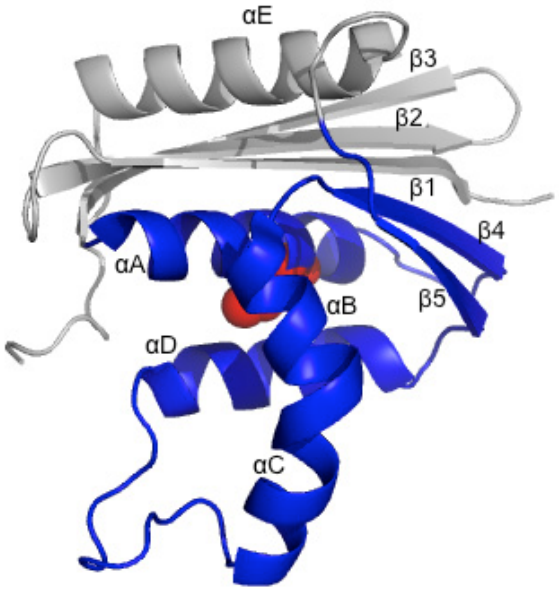


Figure 1.3: Ribbon diagram of *E. coli* RNase H with each secondary structure element named. The region referred to as the core of the molecule is shown in blue. This stretch of the protein was detected as an autonomously folding unit by the rapid autonomous fragment test (RAFT), and it encompasses the regions known to be structured in the kinetic and equilibrium intermediate. The remainder of the protein, referred to as the periphery, is in grey. Isoleucine 53 is shown in red. Upon mutation of this residue to the potentially charged aspartic acid, the protein converts from a three-state folding mechanism to an apparently two-state folding mechanism.

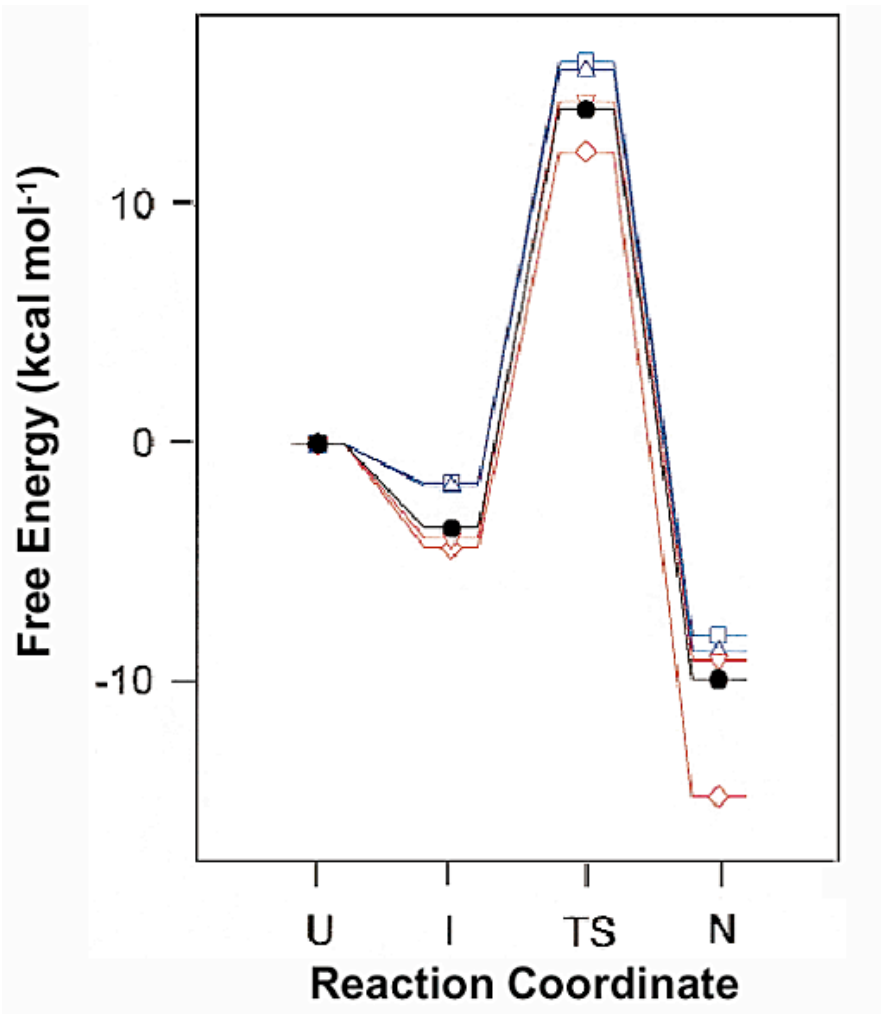
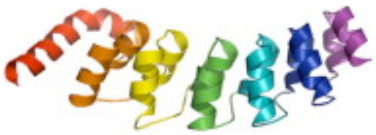
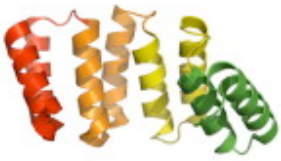


Figure 1.4: Reaction coordinate diagrams for wild type and mutant *E. coli* RNases H in 0 M urea. Wild type is shown as black circles. The destabilizing effects of the core mutations, I53A (blue squares) and Q105G (blue triangles), on the native state are mirrored in the intermediate and transition state. Mutations in the periphery, D10A (red diamonds) and R27A (red triangles), have less of an energetic effect on the intermediate and transition state than on the native state. This figure appears in reference 46.

(a)



(b)



(c)

(d)

(e)

Figure 1.5: Ribbon diagrams of various repeat proteins with adjacent repeats colored red to purple. (a) the Notch ankyrin domain, (b) consensus tetratricorepeats, (c) heat repeats, (d) internalin-B leucine rich repeats, and (e) hexapeptide repeats. This figure appears in reference 95.[95]

Chapter II

The folding trajectory of RNase H is dominated by its topology and not local stability: a protein engineering study of variants that fold via two-state and three-state mechanisms

This chapter is adapted from: Connell, KB et al. (2009). *Journal of Molecular Biology* **391**,450-460.

2.1 ABSTRACT

Proteins can sample a variety of partially folded conformations during the transition between the unfolded and native states. Some proteins never significantly populate these high-energy states and fold by an apparently two-state process. Many proteins, however, populate detectable, partially folded forms during the folding process. The role of such intermediates is a matter of considerable debate. A single amino acid change can convert *E. coli* ribonuclease H from a three-state folder that populates a kinetic intermediate to one that folds in an apparent two-state fashion. We have compared the folding trajectories of the three-state and two-state RNases H, proteins with the same native state topology but altered regional stability, using a protein engineering approach. Our data suggest that that both versions of RNase H fold through a similar trajectory with similar high-energy conformations. Mutations in the core and the periphery of the protein affect similar aspects of folding for both variants, suggesting a common trajectory with folding of the core region followed by the folding of the periphery. Our results suggest that formation of specific partially folded conformations may be a general feature of protein folding that can promote, rather than hinder, efficient folding.

2.2 INTRODUCTION

Elucidating the high-energy conformations a protein samples as it traverses its folding landscape is the key to understanding which factors determine its folding trajectory. Many proteins are known to populate intermediates, and the role of these partially formed species has been the subject of much debate[1]. Intermediates may aid the protein in its search for the native state by greatly reducing the number of accessible conformations[2]. Alternatively, intermediates containing non-native interactions can result in a kinetic trap, inhibiting the folding progress and in some cases leading to the formation of disease-associated amyloid fibrils[3]. What characteristics of these intermediates contribute to their helpful or harmful nature? Furthermore, understanding the nature of the rate-limiting transition state can lend insight into what restrictions are placed on an efficient folding process[4]. How strictly encoded are these species? Can perturbations in these regions of the landscape uncover alternative folding routes to the same native structure?

Several factors are known to guide a protein's folding trajectory. The native state structure, or topology, is clearly important. Contact order, the average sequence separation of native contacts, was the first topological parameter observed to correlate with the free-energy barrier for simple two-state folding proteins[5]. Contact order itself, however, does not appear to be the defining factor[6], and other structural parameters such as long-range order[7], secondary structure content[8], and the total number of long-range contacts[9] have also been observed to correlate with folding rates. Circular permutants of the ribosomal protein S6[10] and α-spectrin SH3[11] fold through different transition states, demonstrating the malleability of the pathway in response to changes in topology. Conversely, comparing the folding trajectories within families of proteins shows that, in many cases, the general properties of the folding pathway are conserved while the details vary widely, highlighting the importance of sequence-dependent effects[12; 13; 14; 15; 16; 17].

While topology clearly plays a role in determining folding pathways, the detailed physical basis of its role is unclear.

Local or regional stability has also been implicated in guiding the folding process and intermediate formation. For the proteins cytochrome c[18] and RNase H,[19] the stability of partially folded forms identified by native-state hydrogen exchange has been shown to correspond with the order of structure formation along the kinetic pathway; the most stable regions appear to fold first. Folding pathways have also been re-routed based on thermodynamic considerations. Tripp and Barrick successfully moved the transition state of an ankyrin repeat protein by introducing stabilized consensus repeats at various locations in the Notch ankyrin domain[20]. These designed repeats were observed to fold first, showing that local energetics guide the folding mechanism of this protein. A similar plasticity was demonstrated with the structurally similar proteins G and L. In the transition state of protein L, the first β hairpin is structured while the second is not, and the opposite is true for the transition state of protein G[21]. The first β hairpin of protein G was redesigned to increase its stability, and the folding of the resulting protein proceeded through the stabilized first turn so that it resembled the folding pathway of protein L[22].

Here, we set out to explore the relative contributions of topology and regional stability by comparing variants of the same protein that fold via distinctly different kinetics. The energy landscape of *E. coli* RNase H has been well-characterized using native-state hydrogen exchange[23; 24; 25], stopped-flow kinetics, quenched-flow hydrogen exchange[19], and protein engineering[26]. The protein populates a kinetic refolding intermediate whose structure is mirrored by a partially folded form detected by the equilibrium native-state hydrogen exchange experiment. Recently, single-molecule experiments confirmed that the intermediate is on-pathway and obligatory[27]. This partially folded form is characterized by a high degree of secondary structure formation in the protein core and a relatively unfolded periphery (Figure 1).

In a point mutant that introduces a potentially-charged residue into this core, I53D, the intermediate is no longer detected, and the protein folds with an apparently two-state mechanism an order of magnitude more slowly than wild type[28]. Using this mutant, we compare the folding mechanisms of a three-state and a two-state version of the protein that differ by only one amino acid. The I53D mutation destabilizes the protein, and the addition of subsequent mutations was expected to result in further destabilization. To counter this effect and ensure sufficient stability of all constructs, we used a stabilized mutant, D10A, as our reference protein[24; 26]. This mutation relieves charge repulsion in the periphery of the protein and stabilizes the protein by more than 3 kcal/mol. Introducing this mutation also allowed us to ask if destabilizing the core and stabilizing the periphery simultaneously (D10A/I53D) results in an altered kinetic trajectory in which folding proceeds through a more diffuse transition state with looser interactions over a greater portion of the protein. A phi-value (ϕ-value) analysis on these variant proteins that share the same RNase H fold, but differ dramatically in their distribution of regional stability, allow us to evaluate the nature of the transition state for both two-state and three-state folders.

2.3 RESULTS

Our reference protein for three-state folding is the variant D10A RNase H, referred to herein as RNH3[24; 26]. The variant D10A/I53D, abbreviated as RNH2, serves

as our two-state reference protein. These two proteins differ only by a single mutation: the insertion of a potentially charged residue in the core of the protein (Figure 1).

Design of Mutants

Significant changes in stability without notable alterations in structure are required to interpret ϕ-values[29]. Since we sought to examine the effect of core mutations on both the three-state and two-state variants, we did not want to destabilize the core of the protein so much that the intermediate would no longer be apparent, as is the case with I53D. This requirement places an upper limit on $\Delta\Delta G_{UN}$ of approximately 3.5 kcal/mol (the stability of the intermediate) for the effect of mutations in this region. A variant containing a mutation in helix D of the core, Q105G (Figure 1a), was previously characterized in the wild type background and is known to destabilize the molecule by just over 2 kcal/mol, making this a good target[26]. The Trp at position 85 in helix C (Figure 1a) was also targeted as a likely residue to destabilize the core. Ile 25 and Phe 8 were chosen as the structural probes for the periphery. Both residues are part of a hydrophobic cluster in this region (Figure 1b).

Equilibrium Characterization of the Mutants

The circular dichroism (CD) spectra of all of the proteins used in this study are shown in Figure 2 (insets). All variants exhibit similar profiles characteristic of the RNase H α/β fold, indicating that the mutations do not significantly disturb the native structure. The stabilities of the variants, as measured by the free energy of unfolding, were determined by urea-induced denaturation profiles assuming a two-state transition and linear extrapolation[30]. A representative normalized denaturation melt of each protein is shown in Figure 2.

Treatment of the equilibrium data. When each denaturation profile was fit independently, variation in the m-value (the denaturant-dependence of the free energy) never exceeded 10 %. Fixing the m-value for all the proteins to that obtained for both the wild type and RNH3 variant (m = 2.1 kcal mol^{-1} M^{-1})[26] fit the data equally well. Since the magnitude of the stability change between the reference proteins and the variants is especially crucial for the calculation of ϕ-values and is highly dependent on the m-value, we calculated the ϕ-values three different ways: 1) using $\Delta\Delta G_{UN}$'s obtained from urea melts with no constraint on the m-value, 2) using $\Delta\Delta G_{UN}$'s from the fits with fixed m-values, and 3) using $\Delta\Delta G_{UN}$'s calculated from fits to the kinetic data (see Tables 1 and 2 for a complete summary). Although absolute values vary somewhat, the general trends and conclusions remain the same. For this reason, we describe the effect of the mutations in terms of the data obtained by fixing the m-values.

Core mutations. The core mutation Q105G was previously shown to destabilize the wild type protein by 2.2 kcal/mol[26]. In the background of RNH3, this mutation has a similar effect, bringing the global stability to 10.5 kcal/mol (compare to 13 kcal/mol for RNH3[26]). In RNH2, this mutation appears to be slightly less destabilizing ($\Delta\Delta G_{UN}$ ~1.9 kcal/mol). The core mutation W85A destabilizes the wild type protein and RNH3 to a similar extent, 2.2 and 2.4 kcal/mol respectively, and the magnitude of this effect also decreases in the background of RNH2, with a change in stability of 1.7 kcal/mol.

Periphery mutations. I25A has an unexpectedly dramatic effect on the stability of RNase H; this variant destabilizes the periphery enough to populate the intermediate under native conditions in the wild type background, prohibiting a full kinetic analysis. In the stabilized background of RNH3, this mutation destabilizes the protein by

approximately 5.1 kcal/mol. In the two-state background, RNH2, I25A destabilizes the protein by 4.4 kcal/mol. The F8A mutation destabilizes the protein by 2.6 kcal/mol in the context of the wild type protein and 3.2 kcal/mol in the background of RNH3. The effect of this mutation decreases to 2.3 kcal/mol in the two-state background of RNH2.

Kinetic phi-value analysis in the two-state and three-state RNases H

Core mutations. The mutations introduced into the core, while destabilizing, were selected to avoid being so disruptive that they abrogate detection of the burst-phase kinetic intermediate. Indeed, a large percentage of the CD signal was acquired in the dead time of the stopped-flow instrument for both Q105G RNH3 and W85A RNH3; these burst-phase signals melted out cooperatively with urea (Figure 3a), as expected for the formation of an early intermediate. Furthermore, the chevron plots display the significant rollover in the folding limbs characteristic of the formation of an intermediate. These data were therefore fit using a three-state mechanism with an on-pathway intermediate in pre-equilibrium with the unfolded state[31]:

$$U \rightleftarrows I \rightarrow N \qquad \text{Scheme 1}$$

The stability of the intermediate was inferred from an analysis of the burst phase amplitudes as a function of urea (Figure 3a)[28]. Both Q105G and W85A shift the C_m of the intermediate to lower urea concentrations. The ratio of the change in stability to the intermediate and the native state, ϕ^I , for Q105G is 0.9 (Table 1), indicating that native-like interactions are formed at this position in the intermediate. W85A, in helix C, has a ϕ^I-value of 0.5, suggesting that there may be some native-like interactions in the intermediate, but that they are only partially formed or are fully formed in only some portion of the molecules in the ensemble.

If these two residues are also structured in the transition state, as is expected, the change in stability should manifest itself in a shift in the folding limb of the chevron while the unfolding limb remains unchanged. This is indeed observed for both Q105G RNH3 and W85A RNH3 (Figure 3a). The quantitative effects of these core mutations corroborate the qualitative observations. The $\phi^{t.s.}$-values, the ratio of the change in stability of the transition state to that of the native state (Equation 1 in Materials and Methods), for both of these mutants are high, 1.3 for Q105G and 0.96 for W85A (Table 1). The $\phi^{t.s.}$-value over 1 for Q105G arises from an unfolding rate that is actually slower than that of the reference protein. Given the errors inherent in the calculation of ϕ-values, we feel it is best to interpret them qualitatively as low, medium, and high rather than quantitatively[32; 33]. $\phi^{t.s.}$-values for these core residues are classified as high, confirming that, as expected, the core region is very important for the energetics of the transition state.

In the background of RNH2, no burst phase intermediate was observed, and there was no apparent rollover in the folding limb of any of the variants (Figure 3b). The chevrons for these proteins were fit to a simple two-state mechanism:

$$U \rightleftarrows N \qquad \text{Scheme 2}$$

These mutations show qualitatively similar shifts in the folding limb of the RNH2 variant as they do in RNH3. Q105G RNH2 and W85A RNH2 have very similar unfolding rates in water as RNH2 but have a steeper unfolding limb, the implications of which are discussed below. The $\phi^{t.s.}$-values (Equation 2 in Materials and Methods) are high (0.95 for Q105G and 1.06 for W85A (Table 2)), similar to those calculated for these

mutations in the background of RNH3. Therefore, in both the three-state and two-state proteins, it appears that the core residues are in a native-like conformation in the rate-limiting transition state.

Periphery Mutations. I25A RNH3 and F8A RNH3 both show a notable burst phase amplitude and rollover in the folding limb of the chevron plots (Figure 3c), and the data were therefore fit to the three-state mechanism shown in Scheme 1. Mutations in the periphery of RNH3 are not expected to affect the energetics of the intermediate or the transition state of the protein, and indeed, in contrast to the core mutations, the C_m for the burst-phase amplitudes in these periphery mutants is similar to that of RNH3 (Figure 3c). The ϕ^I-value for I25A is 0.2 and 0 for F8A (Table 1), confirming that these residues are relatively unstructured in the intermediate. The unfolding limbs of I25A RNH3 and F8A RNH3 are clearly shifted from that of RNH3; these variants unfold more rapidly than the reference protein. The folding limbs also exhibit some shift, however, suggesting that these mutations do not leave the stability of the transition state totally unaffected. The $\phi^{t.s.}$-value (Equation 1) is 0.35 for I25A RNH3 and 0.47 for F8A RNH3 (Table 1). These are intermediate ϕ-values, indicating that there may be some limited native-like structure in this region of the transition state.

In the two-state background, there is no evidence of rollover and no apparent burst phase intermediate (Figure 3d); hence these data were fit to Scheme 2. The trends observed for these mutations in the background of RNH2 are similar to those in RNH3. Dramatic shifts are evident in the unfolding limbs of I25A RNH2 and F8A RNH2 while the folding limbs are slightly shifted. The $\phi^{t.s.}$-values (Equation 2) are also similar to those in the background of RNH3, at 0.25 and 0.27 for I25A RNH2 and F8A RNH2, respectively (Table 2).

Destabilized core/stabilized periphery

We also examined the kinetic trajectory of W85A and F8A in the wild type background (the kinetics of Q105G were reported previously[26], and I25A cannot be examined in the wild type background). This allows a direct comparison between the wild type mechanism and that of RNH2, in which the unfolded part of the intermediate and transition state is stabilized (D10A) and the folded part is destabilized (I53D), a more drastic alteration of the relative stability of the core and periphery.

W85A WT (W85A in the wild type background) shows a rollover in the folding limb along with cooperative melting of burst phase amplitudes (Figure 4a). Although the rollover is less apparent for F8A WT (F8A in the wild type background), this variant also has a large burst-phase amplitude (Figure 4b). Scheme 1 was therefore used to fit the chevron plots of both variants. Qualitatively it is clear from the burst phase signals that the core mutation compromises the stability of the intermediate while the periphery mutation leaves it unaffected, although the ϕ^I-values are not well defined (Table 1). Because the intermediate is only present in very limited range of urea concentrations, the value for m_{UI}, defined by the curvature in the chevron, is ill defined for F8A, and the intermediate appears stabilized, resulting in negative ϕ^I-values. Similarly, the m_{UI} value for W85A is larger than that of wild type, resulting in a calculated value for $\Delta\Delta G_{UI}$ very close to that of wild type. However, it is clear from the burst phase amplitudes that the intermediate is indeed destabilized.

The $\phi^{t.s.}$-values, calculated from the unfolding limb data (Equation 1), appear more reliable. W85A, with a value of 0.8, clearly affects the energetics of the transition

state. F8A shows a moderate $\phi^{t.s.}$-value of 0.5. These are very similar to those calculated in the background of RNH3, although slightly less polarized. It is clear that the transition state of RNH2 is no more diffuse than that of the wild type protein; if anything, structure is more localized to the core.

Burial of solvent accessible surface area in the transition state

The ratio of solvent accessible surface area buried in the transition state compared to the native state, $m_{t.s.}/m_{eq}$, where $m_{t.s.}$ is the change in solvent accessible surface area upon folding and m_{eq} is the equilibrium m-value, is quantified as the Tanford β-value (β_T). β_T was calculated as $1 - (m_{ni}/m_{eq})$ for the three-state proteins, where m_{ni} represents the denaturant dependence of the unfolding rate constant. $m_{t.s.}$ was taken directly as m_{un} for the two-state proteins, where m_{un} is the denaturant dependence of the folding rate constant. For all variants studies, β_T is 0.7 – 0.8, indicating that 70 to 80 % of the total surface area buried upon folding is buried in the transition state (Tables 1 and 2). This species is similar in compactness across all variants.

2.4 DISCUSSION

Both native state topology and regional stability have been shown to play a role in determining protein-folding pathways. To further investigate the relative contributions of these factors, we manipulated the distribution of regional stability in *E. coli* RNase H through mutation while maintaining native state topology. One of these mutations, in the core of the molecule, transforms the three-state folder into one that folds without the accumulation of an intermediate. We used ϕ-value analysis to compare the folding trajectories of the three-state and two-state reference proteins and found that RNase H folds through a structured core and comparatively unstructured periphery regardless.

Coupling of distal regions of RNH2

Each of the four mutations chosen for this investigation destabilized the two-state protein less than the three-state protein. Given that the only difference between these proteins is a single mutation at position 53, this non-additivity suggests that the charge introduced at residue 53 couples to other regions of the protein. Applying a thermodynamic cycle suggests that the magnitude of this coupling effect ranges from 0.6 kcal/mol (residue 105) to 0.9 kcal/mol (residue 8); the value is 0.7 kcal/mol for positions 85 and 25 (Figure 5). Since coupling between the core and the periphery has not been detected in previous experiments[25], this suggests that the coupling is due to the insertion of a charge at position 53 in the background of D10A.

The folding mechanism of RNase H is robust; two-state and three-state RNases H fold through similar pathways

With the introduction of a single mutation into the folding core of RNase H (I53D), the protein is converted from a three-state folder with an observable intermediate to one that folds in an apparently two-state mechanism. A comparison of the $\phi^{t.s.}$–values for each site in each of these backgrounds is presented in Table 3. RNH3 shows high values for the core mutations and intermediate values for the periphery mutations. These intermediate values suggest that although it is less involved, some regions of the periphery may be more important for the structure of the transition state than originally proposed during the initial discovery of this trajectory[19; 26]. It is clear, however, that the transition state of RNH3 is highly polarized with a native-like core and a comparatively unstructured periphery. The periphery of the two-state protein appears to be no more

structured in the transition state than in the three-state protein, with very similar ϕ-values for each mutation. Furthermore, all variants bury a similar proportion of solvent accessible surface area upon folding.

The wild type trajectory can also be compared to that of RNH2, which has two residue changes resulting in a stabilized periphery as well as a destabilized core. Even these drastic changes in the distribution of stability do not appear to re-route folding. This is somewhat surprising considering that folding of the wild type enzyme appears to follow a hierarchical mechanism in which the most stable regions fold first. When the core is destabilized by more than 6 kcal/mol, in both the I53D/W85A and I53D /Q105G mutations, this region is still capable of initiating folding without apparently incorporating additional elements from the periphery. The stability of the core of the molecule only determines whether the protein folds through an observable intermediate in a three-state fashion or through an apparent two-state mechanism. When this region is destabilized to such an extent that this intermediate is no longer detectable, this region still dominates the folding trajectory and presumably the intermediate species still exists, albeit at a higher energy on the folding trajectory[28]. Apparently, the topology of RNase H places strict constraints on the order of folding events and gives rise to the apparent robustness of this mechanism.

What makes RNase H different from proteins such as protein G and the Notch ankyrin domain, whose pathways were successfully re-routed based on thermodynamic considerations? One possible explanation is the simplified architecture of these two proteins. The notch protein is made up of repeats of the ankyrin domain, consisting entirely of sequence-local interactions. The modular arrangement of structurally identical units may account for its plasticity in response to changes in the distribution of stability. Proteins G and L are small (~60 residues) and symmetric in structure[34]. Though globular, these proteins can be thought of as modular as well, with two structurally identical halves, in which folding proceeds through the half with the more stable β-hairpin.

Changes in m_{nu} for RNH2 mutants

It is interesting to note that in the background of RNH2, the values for m_{nu} of the core mutants increases quite drastically while they decrease very slightly for the periphery mutants. The slight decrease for I25A and F8A RNH2 results in kinetic m-values that agree nicely with the equilibrium m-values. This is not the case for the core mutants Q105G and W85A RNH2, whose kinetic m-values are greater than those calculated from the fit to the equilibrium denaturant melts. The increase in slope of the unfolding limb for the core mutants can also be detected in the background of RNH3, though to a much lesser extent.

Increased values for m_{nu} upon mutation can arise from a movement of the transition state, a complete change in the barrier traversed upon folding, or ground state effects in the native state[35]. The equilibrium m-values for all proteins studied are the same within error, and ground state effects are much less common for the native state than the unfolded state. Whether this increase in m_{nu} reflects a complete change in mechanism under high urea or movement[36] of the transition state is unclear and in a sense describes two extremes of the same phenomenon. This debate is beyond the scope of this paper. Nonetheless, ϕ-values calculated from the unfolding rates in 8M urea are ~0.5 as opposed to ~1, suggesting that in high urea, the protein may indeed not incorporate as much of the core into the transition state. Under conditions in which the native state is

destabilized, movement of the structure of the transition state away from that of the native is indicative of anti-Hammond behavior. Similar increases in the m-value for unfolding upon mutation have been observed and interpreted as anti-Hammond behavior for TI I27[37]. One explanation for the observed shift in m_{nu} in our two state proteins is that in high levels of denaturant and with the folding core severely compromised, folding can proceed through a more loosely structured core.

Ruggedness in the energy landscape can be productive for folding

The results presented here support the argument that formation of specific ensembles of partially-folded species are a universal feature of folding that play an important role even for proteins that appear to proceed in a two-state manner. The development of new tools to study the progression of folding may uncover these important high-energy ensembles. For instance, novel NMR studies have recently uncovered a high energy intermediate for the FynSH3 domain, which was until recently considered to be a two-state system[36]. Trapping partially folded intermediates at equilibrium to gain experimental access is another novel approach. This can be accomplished by introducing destabilizing mutations in the unfolded regions of the intermediate[38; 39; 40] (also see Connell, Horner, and Marqusee, accompanying manuscript) or by constructing a fragment representing the folded region.[41]

Folding studies between different members of a protein family also suggest that two-state behavior masks some of the inherent complexities of folding. Both two-state and multi-state folding has been observed in the immunity proteins[13], immunoglobulin-like proteins[37], and members of the homeodomain family[42; 43]. Im7 was observed to populate a kinetic intermediate while Im9, sharing 60% sequence identity, appeared two-state[13]; however, under different conditions both proteins folded through an intermediate, albeit with different structural properties[17]. Changing just a few key residues of Im9 led to the population of an intermediate closely resembling that of Im7 under the original conditions in which Im9 appeared two-state[14]. Changes in sequence or in folding conditions have actually led to the switching from one kinetic mechanism to another for members of all of the above families, suggesting that high-energy structures are only detectable with increases in the ruggedness of the landscape large enough to bring the stability of the species below that of other states.

It is important to note that this increased ruggedness does not necessarily slow folding. In the case of RNase H, populating the intermediate actually promotes efficient folding; the folding rate of RNH3 is an order of magnitude higher than that of RNH2. Wagner and Kiefhaber[44] explored the effect of intermediates on the rate of protein folding applying Kramer's theory of diffusive barrier and found that introducing additional local barriers and valleys (roughness) can actually enhance the folding rate without altering the rate-limiting barrier. In light of the possibility that such partially folded ensembles are ubiquitous and that transient population of such species can promote rather than hinder folding, their characterization remains a crucial step in enhancing our understanding of folding.

2.5 MATERIALS AND METHODS

Construction and purification of RNase H variants

The RNase H variants described in this manuscript were generated using the Quikchange mutagenesis protocol. A plasmid containing the gene encoding D10A

RNase H[26] was used as a template to create the double mutants. These variants were expressed and purified as described[19]. The triple mutants were created using a plasmid containing D10A/I53D RNase H, pGH101, as a template. These variants expressed insolubly as inclusion bodies and were purified accordingly[28].

CD Spectra

CD spectra were obtained on an Aviv 410 circular dichroism spectropolarimeter. Protein concentrations were determined based on the extinction coefficient, calculated according to the number of Trp and Tyr residues[45]. The buffer conditions for all experiments were 20 mM sodium acetate pH 5.5 and 50 mM potassium chloride. Spectra were collected with a 1 mm pathlength cuvette at a protein concentration of 500 μg/mL.

Equilibrium urea denaturation

For each variant described, samples containing varying amounts of urea and 50 μg/mL protein were equilibrated for the required time and the CD signal was measured at 222 nm at 25°C in a 1 cm cuvette. The signal over 60 seconds was averaged for each sample. The slow folding and unfolding of many of the variants used in this study necessitates long equilibration times, over a week for the slowest folding proteins. All equilibrium denaturation melts were repeated at least three times. The data were fit to a two-state model with a linear free energy extrapolation[30]. Even though the kinetics of the three-state proteins reveal a transient intermediate, at equilibrium this species never populates to more than a few percent due to its substantially lower stability compared to the native state. The equilibrium curves of the three-state proteins can be well-represented by a two-state model, as the unfolded and native forms are the dominant species.

Kinetics of folding and unfolding

Refolding of the mutants was initiated in one of two ways. For the double mutant in low denaturant conditions, refolding was initiated by a 1:11 dilution from a urea concentration in the unfolded baseline and monitored using an Aviv 202-SF stopped-flow device (1 mm pathlength). Final protein concentration in the stopped-flow device was at least 500 μg/mL. For the double mutants in higher urea conditions and for all triple mutants, refolding was induced by a 1:30 dilution via manual mixing and monitored using an Aviv 410 spectrometer. Unfolding for all variants was initiated by manual mixing and a 1:30 dilution from a urea concentration in the folded baseline. The final protein concentration for all manual mixing experiments was 50 μg/mL. Progression of folding and unfolding was monitored at 222 nm, and the resulting curves were fit to a single exponential. Kinetic data were fit as previously described[28].

Calculation of ϕ-values

ϕ-values in the folding direction for the three-state variants were calculated as

$$\phi^{t.s.} = 1 - \frac{RT\ln(k_{ni}^{mut}/k_{ni}^{wt})}{\Delta G_{mut} - \Delta G_{wt}} \qquad \text{Equation 1}$$

$$\phi^{t.s.} = \frac{RT\ln(k_{f}^{mut}/k_{f}^{wt})}{\Delta G_{wt} - \Delta G_{mut}} \qquad \text{Equation 2}$$

For the two-state proteins, ϕ-values were calculated directly from the folding rates:

2.6 ACKNOWLEDGEMENTS

We thank the members of the Marqusee laboratory for helpful discussions and Rachel Bernstein for thoughtful comments on the manuscript. This work was supported by NIH grant GM50945 to S.M. and an NSF graduate research fellowship to K.B.C.

2.7 REFERENCES

1. Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-61.
2. Brockwell, D. J. & Radford, S. E. (2007). Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* **17**, 30-7.
3. Jahn, T. R. & Radford, S. E. (2005). The Yin and Yang of protein folding. *FEBS J* **272**, 5962-70.
4. Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122-6.
5. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985-94.
6. Miller, E. J., Fischer, K. F. & Marqusee, S. (2002). Experimental evaluation of topological parameters determining protein-folding rates. *Proc Natl Acad Sci U S A* **99**, 10359-63.
7. Gromiha, M. M. & Selvaraj, S. (2001). Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J Mol Biol* **310**, 27-32.
8. Gong, H., Isom, D. G., Srinivasan, R. & Rose, G. D. (2003). Local secondary structure content predicts folding rates for simple, two-state proteins. *J Mol Biol* **327**, 1149-54.
9. Makarov, D. E. & Plaxco, K. W. (2003). The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. *Protein Sci* **12**, 17-26.
10. Lindberg, M., Tangrot, J. & Oliveberg, M. (2002). Complete change of the protein folding transition state upon circular permutation. *Nat Struct Biol* **9**, 818-22.
11. Viguera, A. R., Serrano, L. & Wilmanns, M. (1996). Different folding transition states may result in the same native structure. *Nat Struct Biol* **3**, 874-80.
12. Dalessio, P. M. & Ropson, I. J. (2000). Beta-sheet proteins with nearly identical structures have different folding intermediates. *Biochemistry* **39**, 860-71.
13. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **286**, 1597-608.
14. Friel, C. T., Beddard, G. S. & Radford, S. E. (2004). Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J Mol Biol* **342**, 261-73.
15. Geierhaas, C. D., Best, R. B., Paci, E., Vendruscolo, M. & Clarke, J. (2006). Structural comparison of the two alternative transition states for folding of TI I27. *Biophys J* **91**, 263-75.
16. Geierhaas, C. D., Paci, E., Vendruscolo, M. & Clarke, J. (2004). Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *J Mol Biol* **343**, 1111-23.

17. Gorski, S. A., Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2001). Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. *J Mol Biol* **312**, 849-63.
18. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192-7.
19. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* **4**, 298-304.
20. Tripp, K. W. & Barrick, D. (2008). Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape. *J Am Chem Soc* **130**, 5681-8.
21. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* **7**, 669-73.
22. Nauli, S., Kuhlman, B. & Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nat Struct Biol* **8**, 602-5.
23. Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* **3**, 782-7.
24. Goedken, E. R. & Marqusee, S. (2001). Native-state energetics of a thermostabilized variant of ribonuclease HI. *J Mol Biol* **314**, 863-71.
25. Spudich, G., Lorenz, S. & Marqusee, S. (2002). Propagation of a single destabilizing mutation throughout the Escherichia coli ribonuclease HI native state. *Protein Sci* **11**, 522-8.
26. Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* **6**, 825-31.
27. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057-60.
28. Spudich, G. M., Miller, E. J. & Marqusee, S. (2004). Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol* **335**, 609-18.
29. Sanchez, I. E. & Kiefhaber, T. (2003). Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J Mol Biol* **334**, 1077-85.
30. Pace, C. N. & Shaw, K. L. (2000). Linear extrapolation method of analyzing solvent denaturation curves. *Proteins* **Suppl 4**, 1-7.
31. Baldwin, R. L. (1996). On-pathway versus off-pathway folding intermediates. *Fold Des* **1**, R1-8.
32. Fersht, A. R. & Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci U S A* **101**, 7976-81.
33. Geierhaas, C. D., Salvatella, X., Clarke, J. & Vendruscolo, M. (2008). Characterisation of transition state structures for protein folding using 'high', 'medium' and 'low' {Phi}-values. *Protein Eng Des Sel* **21**, 215-22.
34. Kim, D. E., Fisher, C. & Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. *J Mol Biol* **298**, 971-84.
35. Sanchez, I. E. & Kiefhaber, T. (2003). Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J Mol Biol* **327**, 867-84.

36. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586-90.
37. Fowler, S. B. & Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure* **9**, 355-66.
38. Feng, H., Vu, N. D. & Bai, Y. (2004). Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome b562: native fold with non-native hydrophobic interactions. *J Mol Biol* **343**, 1477-85.
39. Feng, H., Zhou, Z. & Bai, Y. (2005). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc Natl Acad Sci U S A* **102**, 5026-31.
40. Spence, G. R., Capaldi, A. P. & Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**, 215-26.
41. Zhou, Z., Feng, H., Ghirlando, R. & Bai, Y. (2008). The high-resolution NMR structure of the early folding intermediate of the Thermus thermophilus ribonuclease H. *J Mol Biol* **384**, 531-9.
42. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W., DeMarco, M. L., Daggett, V. & Fersht, A. R. (2003). Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A* **100**, 13286-91.
43. White, G. W., Gianni, S., Grossmann, J. G., Jemth, P., Fersht, A. R. & Daggett, V. (2005). Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J Mol Biol* **350**, 757-75.
44. Wagner, C. & Kiefhaber, T. (1999). Intermediates can accelerate protein folding. *Proc Natl Acad Sci U S A* **96**, 6716-21.
45. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948-54.

Table 2.1: Summary of equilibrium and kinetic data for the three-state RNases H. The values in parentheses indicate the standard deviation calculated from at least three separate denaturant melts.

			Periphery		Core		Periphery	Core
		RNH3*	**I25A RNH3**	**F8A RNH3**	**Q105G RNH3**	**W85A RNH3**	**F8A**	**W85A**
$\Delta G_{urea\ melt}$ (kcal mol^{-1})		13	8.2 (0.4)	10.0 (0.2)	10.6 (0.5)	11.3 (0.5)	6.5 (0.5)	6.8 (0.5)
$m_{eq\ urea\ melt}$ (kcal mol^{-1} M^{-1})		2.1	2.2 (0.1)	2.14 (0.03)	2.11 (0.05)	2.24 (0.1)	1.9 (0.2)	1.9 (0.1)
$\Delta G_{fixed\ m}$ (kcal mol^{-1})			7.86 (0.07)	9.78 (0.07)	10.5 (0.2)	10.56 (0.06)	7.05 (0.08)	7.5 (0.1)
$\Delta G_{kinetic}$ (kcal mol^{-1})		13.4	8.1	11.02	11.9	11.67	9.22	8.6
$m_{kinetic}$ (kcal mol^{-1} M^{-1})		2.25	2.13	2.35	2.43	2.37	2.7	2.48
k_{in} (H_2O) (s^{-1})		1.89	0.37	0.55	0.59	0.48	0.227	0.1514
m_{in} (kcal mol^{-1} M^{-1})		0.28	0.49	0.50	0.15	0.19	0.647	0.415
k_{ni} (H_2O) (s^{-1})		1.5E-07	4.4E-05	2.6E-06	5.5E-08	1.6E-07	9.50E-05	2.50E-05
m_{ni} (kcal mol^{-1} M^{-1})		-0.48	-0.49	-0.48	-0.60	-0.51	-0.529	-0.472
K_{ui}		585	110	611	54	133	2580	350
m_{ui} (kcal mol^{-1} M^{-1})		1.50	1.15	1.37	1.68	1.66	1.505	1.596
ΔG_{ui} (kcal mol^{-1})		3.76	2.77	3.79	2.35	2.89	4.64	3.46
ϕ^{I}	1		0.19	-0.01	0.56	0.36	-0.42	0.04
	2		0.2	0.0	0.6	0.5	-0.1	0.0
	3		0.19	-0.01	0.94	0.51	-1.64	0.07
$\phi^{t.s.}$	1		0.35	0.47	1.23	0.97	0.51	0.78
	2		0.3	0.4	1.2	1.0	0.4	0.5
	3		0.37	0.29	1.39	0.96	-0.93	0.62
β_T	1	0.77	0.77	0.77	0.71	0.76	0.75	0.78
	2	0.8	0.8	0.8	0.7	0.8	0.7	0.8
	3	0.79	0.77	0.80	0.75	0.78	0.80	0.81

*These values were previously determined by Raschke, Kho, and Marqusee[26].

1. calculated using a fixed m-value of 2.1
2. calculated using the m-value from urea melts
3. calculated using the m-value from the kinetic data; $m_{kinetic} = m_{ui} + m_{in} - m_{ni}$

Table 2.2: Summary of equilibrium and kinetic data for the two-state RNases H. The values in parentheses indicate the standard deviation calculated from at least three separate denaturant melts.

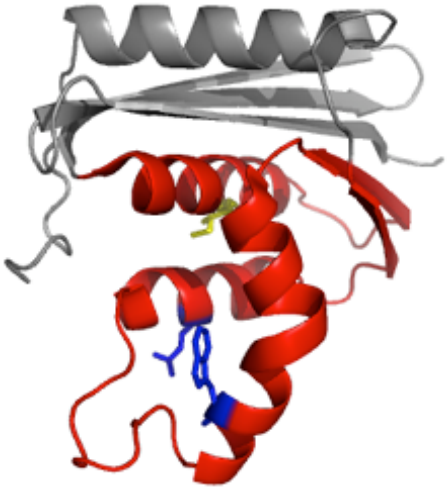
			Periphery		Core	
		RNH2	**I25A RNH2**	**F8A RNH2**	**Q105G RNH2**	**W85A RNH2**
$\Delta G_{urea\ melt}$ (kcal mol^{-1})		8.5 (0.4)	4.8 (0.1)	6.32 (0.4)	7.1 (0.5)	7.2 (0.3)
$m_{eq\ urea\ melt}$ (kcal mol^{-1} M^{-1})		2.0 (0.1)	2.24 (0.01)	2.0 (0.1)	2.1 (0.1)	2.08
$\Delta G_{fixed\ m}$ (kcal mol^{-1})		8.9 (0.1)	4.5 (0.1)	6.62 (0.08)	7.05 (0.08)	7.219 (0.006)
$\Delta G_{kinetic}$ (kcal mol^{-1})		9.09	4.60	6.56	7.24	7.76
$m_{kinetic}$ (kcal mol^{-1} M^{-1})		2.07	2.05	2.13	2.31	2.34
k_{un} (H_2O) (s^{-1})		0.40	0.058	0.126	0.020	0.037
m_{un} (kcal mol^{-1} M^{-1})		1.51	1.58	1.64	1.62	1.67
k_{nu} (H_2O) (s^{-1})		8.10E-08	2.40E-05	2.10E-06	9.54E-08	7.61E-08
m_{nu} (kcal mol^{-1} M^{-1})		-0.56	-0.47	-0.49	-0.69	-0.67
$\phi^{t.s.}$	1		0.26	0.30	0.95	0.84
	2		0.31	0.31	1.26	1.05
	3		0.25	0.27	0.95	1.06
β_T	1	0.72	0.75	0.78	0.77	0.80
	2	0.76	0.71	0.80	0.77	0.80
	3	0.73	0.77	0.77	0.70	0.71

1. calculated using a fixed m-value of 2.1
2. calculated using the m-value from urea melts
3. calculated using the m-value from the kinetic data

Table 2.3: Direct comparison of $\phi^{t.s.}$-values for the four mutations in each of the three backgrounds.

	$\phi^{t.s.}$			
	Periphery		Core	
	I25A	**F8A**	**Q105G**	**W85A**
Three-State	0.35	0.47	1.23	0.97
Two-State	0.31	0.30	0.95	0.84
wt		0.51	1.19	0.78

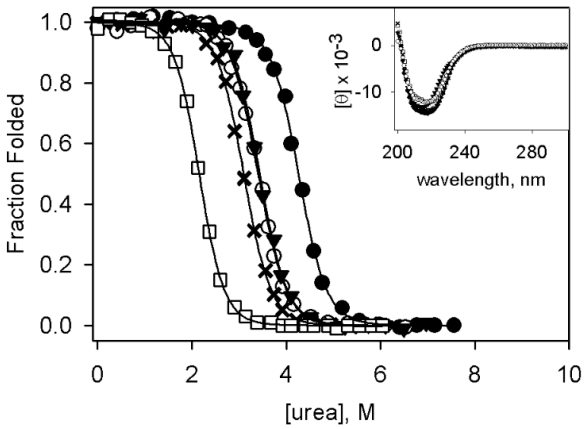
(a)



(b)

Figure 2.1: Structure of E. coli RNase H with the "core" colored red and "periphery" in grey. (a) The core mutations are shown as sticks, with Ile 53 in yellow and Asp 105 and Trp 85 in blue. (b) The periphery mutations, Phe 8 and Ile 25, are shown in green.

(a)



(b)

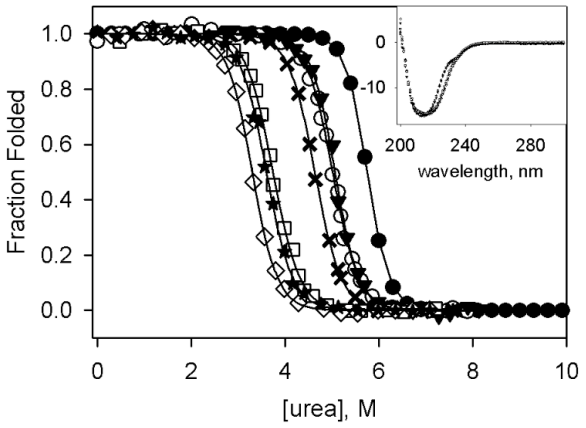
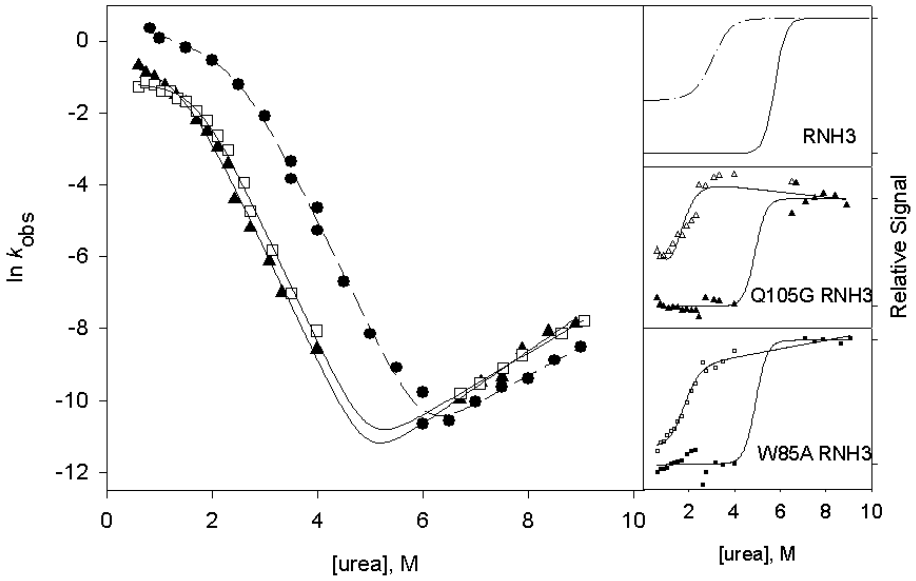
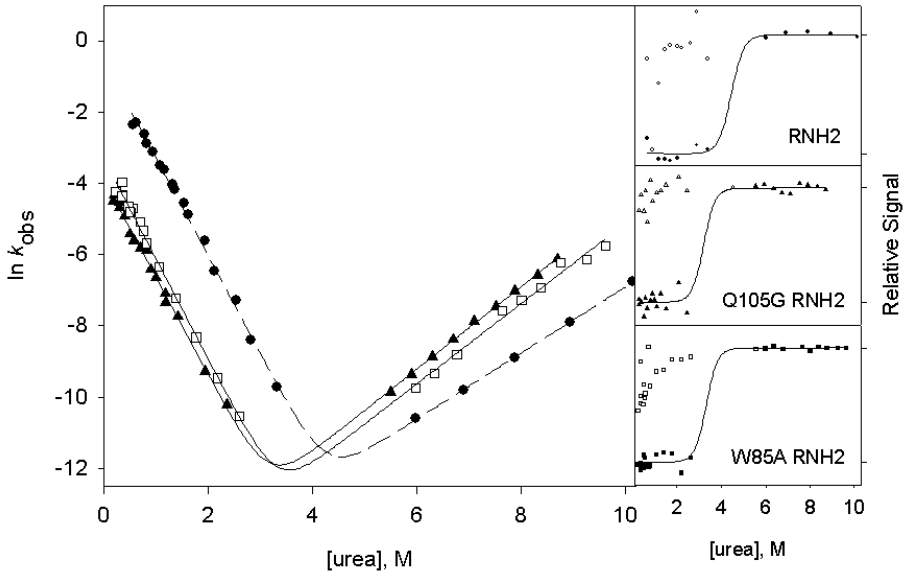


Figure 2.2: Representative urea denaturation curves of variants of RNase H normalized to fraction folded. The two-state (a) and three-state (b) proteins are separated for clarity. Black lines represent the linear extrapolation fit assuming a two-state model. The data for the reference proteins RNH2 in panel (a) and RNH3 in panel (b) are shown as filled circles, and the mutations in each of these backgrounds are as follows: Q105G (open circles), W85A (closed triangles), F8A (filled crosses), I25A (open squares). In panel (b) F8A in the wild type background is shown as open diamonds, and W85A in the wild type background is shown as stars. CD spectra of the mutants are shown as insets in the same symbols.

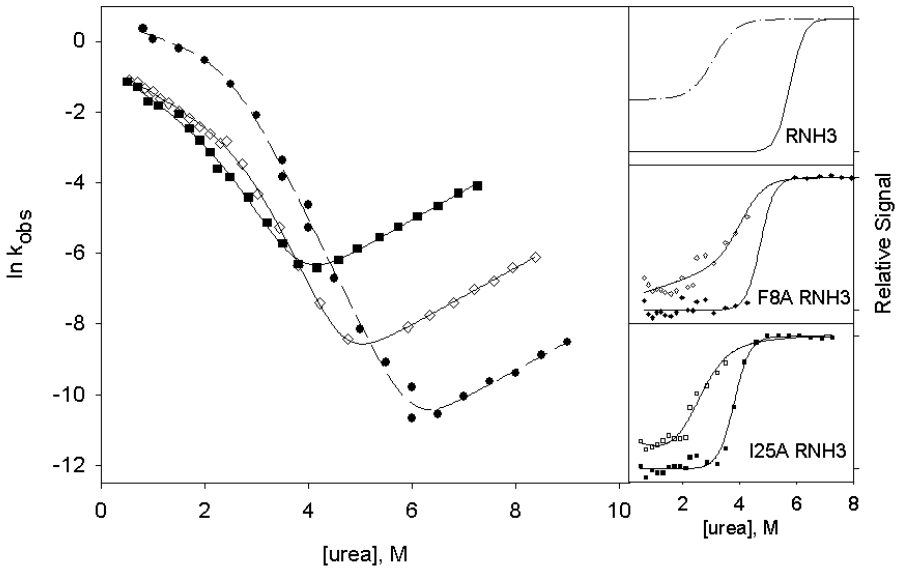
(a)


ln k_{obs}
0
-2
-4
-6
-8
-10
-12
0
2
4
6
8
10
[urea], M
RNH3
Q105G RNH3
W85A RNH3
Relative Signal
2
4
6
8
10
[urea], M

(b)


ln k_{obs}
0
-2
-4
-6
-8
-10
-12
0
2
4
6
8
10
[urea], M
RNH2
Q105G RNH2
W85A RNH2
Relative Signal
2
4
6
8
10
[urea], M

(c)



(d)

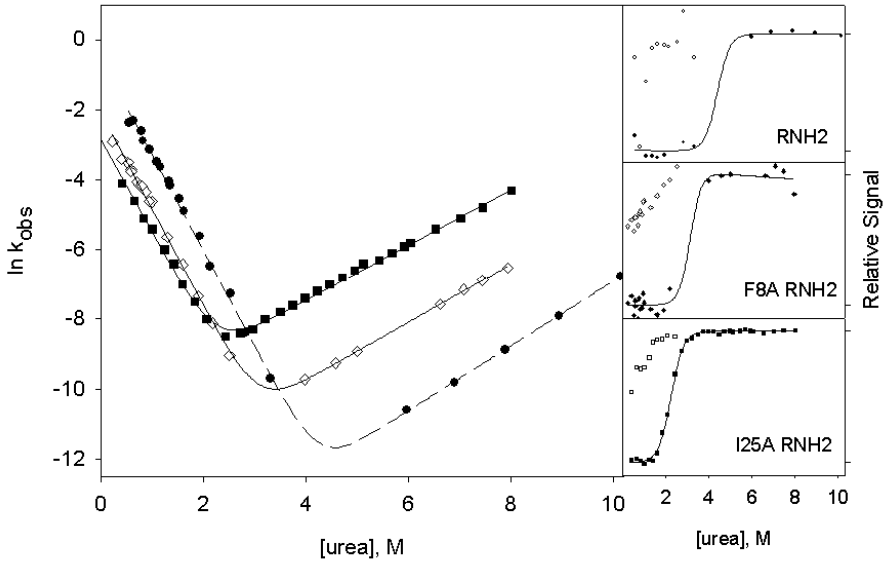


Figure 2.3: Chevron plots for the three-state ((a) and (c)) and two-state ((b) and (d)) RNases H. The data for the reference proteins, RNH3 and RNH2, are filled circles. Panels (a) and (b) show the core mutation Q105G (filled triangles) and W85A (open

squares). Panels (c) and (d) show the data for the periphery mutations F8A (open diamonds) and I25A (filled squares). The solid lines in (a) and (c) represent the fit to Scheme 1 and in (b) and (d) the fit to Scheme 2. The amplitudes are shown to the right of each chevron plot, with the initial signal shown as open symbols and final shown in closed symbols. The data are plotted as the fraction of relative overall signal. The black lines represent a linear extrapolation fit obtained by fixing the value for ΔG and m calculated from the chevron fit to show that they are consistent.

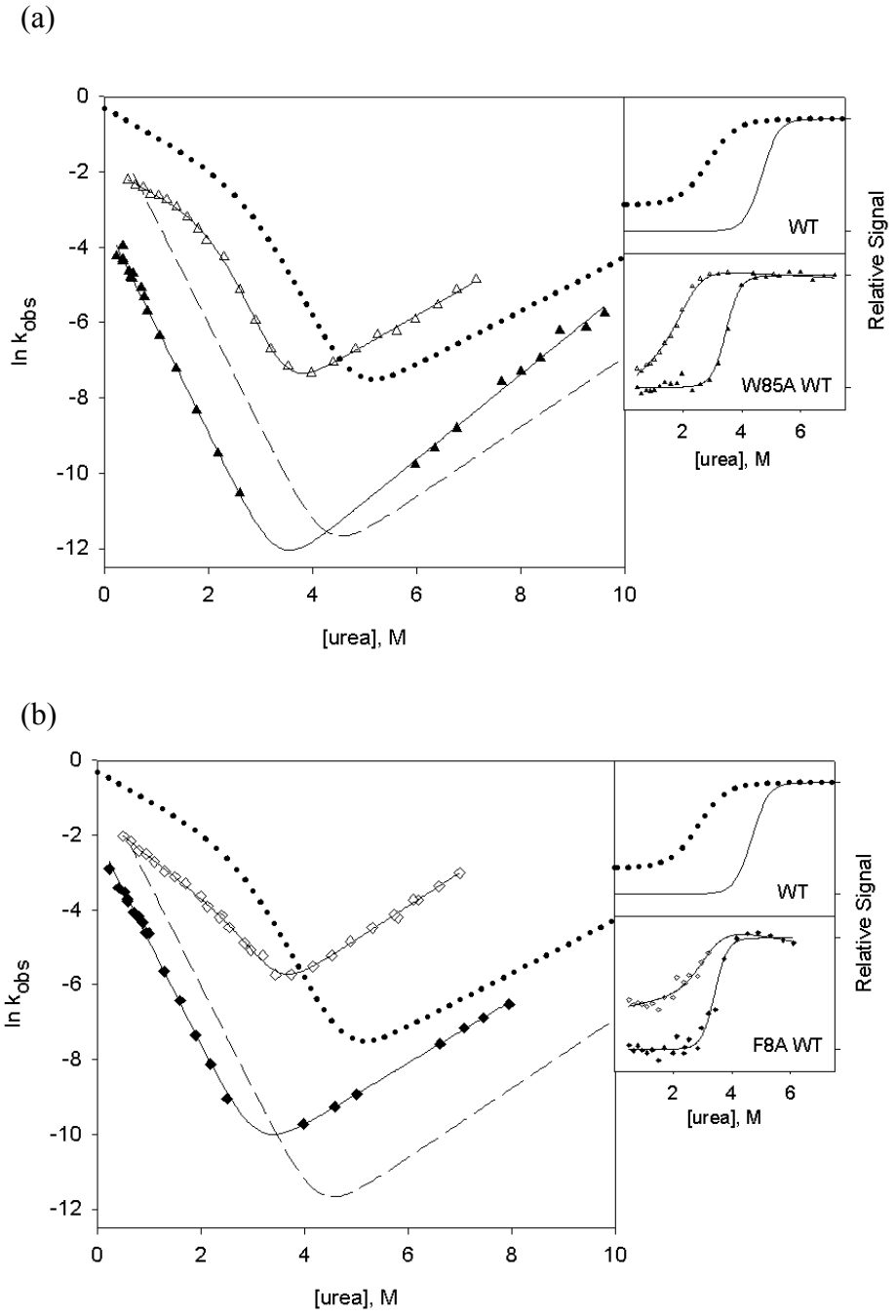


Figure 2.4: Chevron plots for the core mutation W85A (a) and the periphery mutation F8A (b) in the wild type background. W85A RNH3 is shown as open triangles and W85A RNH2 as closed triangles. The dotted lines represent the wild type chevron while the dashed line represents a fit to the RNH2 chevron. F8A RNH3 is shown as open diamonds and F8A RNH2 as closed diamonds. The data for the two-state proteins are shown as in Figure 3 for comparison. The amplitudes are shown as in Figure 3.

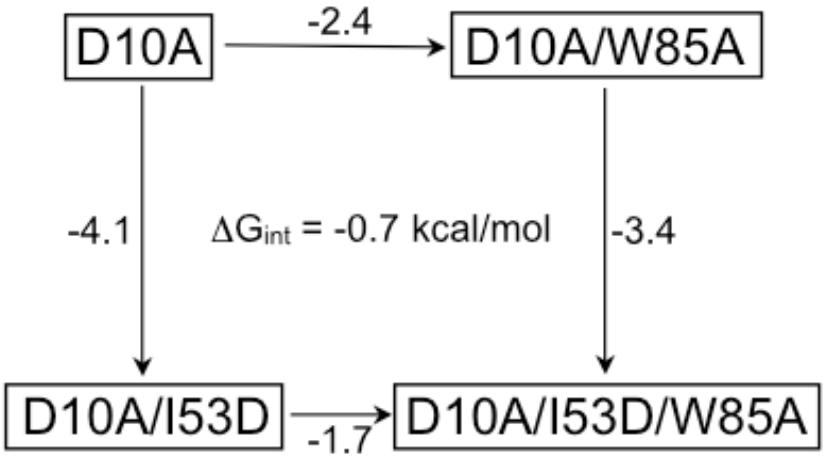


Figure 2.5: A thermodynamic cycle addressing the interaction between Ile 53 and the mutated sites, represented here by Trp 85. The ΔΔG (kcal/mol) of unfolding upon making the indicated mutation is reported along the arrows. This cycle shows an interaction energy of 0.7 kcal/mol for residues 53 and 85.

Chapter III

A single mutation at residue 25 populates the folding intermediate of *E. coli* RNase H and reveals a highly dynamic partially-folded ensemble

This chapter is adapted from: Connell, KB et al. (2009). *Journal of Molecular Biology* **391**, 461-470.

3.1 ABSTRACT

Understanding the nature of partially folded intermediates transiently populated during protein folding is important for understanding both protein folding and misfolding. These ephemeral species, however, often elude direct experimental characterization. The well-characterized protein ribonuclease H from E. coli populates an on-pathway intermediate identified in both bulk studies and single molecule mechanical unfolding experiments. Here, we set out to trap the transient intermediate of RNase H at equilibrium by selectively destabilizing the region of the protein known to be unfolded in this species. Surprisingly, a single change at Ile 25 (I25A) resulted in the equilibrium population of the intermediate under near-native conditions. The intermediate was undetectable in a series of HSQC spectra, revealing the dynamic nature of this partially unfolded form on the timescale of NMR detection. This result is in contrast to studies in which the structures of trapped intermediates are solved by NMR, indicating that the they are well-packed and native-like. The dynamic nature of the RNase H intermediate may be important for its role as an on-pathway, productive species that promotes efficient folding.

3.2 INTRODUCTION

Partially folded intermediates are known to play an important role in the mechanism of protein folding. For some proteins, intermediates appear to play a productive role, aiding in the formation of the native fold, while for others they constitute an off-pathway species. In addition to their importance in folding mechanisms, partially folded forms may also play crucial biological roles. There are many examples of proteins with unstructured regions whose disordered-ordered folding transitions are important for binding and other regulatory events[1;2]. For example, the active form of the steroidogenic acute regulatory protein appears to be a molten globule[3], a state usually associated with early folding intermediates. Such intermediates have also been implicated in the formation of aggregation-prone species[4]. In spite of their clear importance in both folding and function, these partially folded intermediates are usually not amenable to detailed structural and biophysical studies due to their low populations at equilibrium and transient nature.

Here, we set out to trap the intermediate of ribonuclease H from *E. coli* (RNase H), a small (17.5 kDa), monomeric enzyme known to fold through such a partially folded kinetic intermediate[5; 6; 7] (throughout this manuscript, wild type RNase H refers to the *E. coli* variant with all three cysteines replaced by alanine). Recent single molecule experiments conclusively demonstrated that this species is on-pathway and obligatory[8]. Characterization of the intermediate by phi-value analysis and quenched flow hydrogen exchange suggest that the intermediate is composed of native-like topology in the core and a relatively unstructured periphery (Figure 1a)[5; 6]. This partially folded ensemble appears to play a robust and important role in the folding trajectory of RNase H, dominating the folding trajectory even in variants that do not transiently populate the state (see accompanying manuscript). Equilibrium native state hydrogen exchange experiments detect a rare high energy partially unfolded form whose structure and energetics mirror this kinetic intermediate, suggesting that the most stable region of the protein is the first to fold[6; 9].

We have used rational design to selectively destabilize the native state of the protein, enriching the population of the high-energy species. The response of RNase H to mutations highlights the differential distribution of stability in the core and periphery that makes this approach possible. Isoleucine 53, located in the core, was mutated to alanine (I53A), destabilizing the protein by ~2 kcal/mol ($\Delta\Delta G_{UN}$ ~ 2kcal mol^{-1})[10]. Native state hydrogen exchange demonstrated that this destabilizing effect was localized to the core, while residues in the periphery were unaffected, resulting in destabilization of the intermediate. A similar study was carried out in the periphery with the stabilizing mutation D10A[11], where, although all secondary elements showed an increase in the free energy of unfolding as measured by native state hydrogen exchange, the difference between unfolding the periphery and the core remained unchanged, suggesting that the two regions are energetically independent. This selective increase in stability causes a decrease in the population of the partially folded form. For both of these variants, the effects in the equilibrium energetics were mirrored in the transient intermediate observed during the folding process. In fact, a severely destabilizing mutation in the core of the protein (I53D) destabilizes the intermediate such that the protein folds in an apparent two-state manner without the obvious accumulation of the intermediate[7].

In principle, then, we should be able to use mutations that destabilize the periphery of the protein without perturbing the stability of the intermediate to selectively destabilize the native state, enhancing the population of the high-energy intermediate such that we can study it more easily. This strategy has been demonstrated by Bai[12; 13; 14; 15; 16; 17] who, based on native state hydrogen exchange data, create mutations to selectively destabilize the regions predicted to be unfolded in the intermediate, thereby selectively destabilizing the native state. Similarly, Radford and coworkers destabilized regions of the immunity protein Im7 based on results from phi-value analysis (monitoring the folding kinetics for engineered site-specific mutations)[18].

In an alternative approach, Bai and coworkers also recently designed a fragment of *T. thermophilus* RNase H modeled, in part, on our native state hydrogen exchange data. By removing two strands believed to be disordered in the intermediate, they obtained a well-folded species amenable to high resolution NMR spectroscopy. Their results reveal a fragment that folds into a well-packed native-like species, suggesting that the structure of the partially folded intermediate may be a subset of the native structure. These results are at odds with studies that suggest the intermediate state of RNase H is a heterogeneous ensemble with the hallmarks of a molten globule[19; 20]. While the fragment approach has the advantage of creating a fragment molecule amenable for high-resolution studies, there are drawbacks. The subjective assumptions about the structure, based on the limited hydrogen exchange data, may omit important structural components of the intermediate. All interpretations are based on the underlying assumption that the fragment is indeed a representative mimic of the high-energy ensemble. Questions therefore remain as to the exact nature of the contacts formed in the high-energy intermediate of the protein, the extent to which tertiary structure has been fixed and the breadth of this species, and which regions contribute to the 'molten' nature of the intermediate.

Here, we set out to trap the intermediate of RNase H by rational design of destabilizing mutants believed to selectively affect native state stability. We make no assumptions as to the structure of the species, but instead use mutagenesis to selectively

destabilize the native state to enrich the population of the high-energy species. This approach is completely analogous to Bai's protein engineering approach to uncover hidden intermediates, only in our case, the intermediate is not hidden, but rather a detectable transiently populated species.

To our surprise, the first target for mutation, isoleucine 25, located in Strand II of the periphery and completely buried in a hydrophobic pocket (Figure 1a), was almost sufficient to accomplish our task. This variant, which is part of a series of mutations used to investigate the robustness of the folding trajectory in two-state and three-state versions of the protein (Connell et al., accompanying manuscript), results in a protein that detectably populates the kinetic intermediate at equilibrium. Although we could not obtain high-resolution structural information, with this mutation we were able to gain valuable insight into the dynamic properties of the intermediate.

3.3 RESULTS

I25A populates the intermediate

Equilibrium urea denaturation of I25A was monitored by circular dichroism (CD) at 222 nm and fit with the standard two-state linear extrapolation method (Figure 1b)[21]. This analysis results in surprisingly low values for both ΔG and the m-value (the denaturant dependence of ΔG_{UN}). The calculated ΔG_{UN} is 3.4 kcal/mol, indicating an abnormally large destabilization from the wild type protein (WT ΔG_{UN} = 9.7 kcal/mol, see Table 1), and the m-value is 1.3 kcal mol^{-1} M^{-1}, much lower than the wild type m-value of 2.1 kcal mol^{-1} M^{-1} (Table 1). The m-value is believed to correlate with the burial of surface area upon folding[22]; therefore, if the overall fold is retained, this parameter should remain unchanged. The m-value we obtained for I25A, while clearly different from that expected based on the native structure, is strikingly similar to that determined for the transient folding intermediate of RNase H[5], suggesting that the native state was unexpectedly destabilized to the extent that the intermediate is now the most stable species.

To further investigate the possibility that the I25A variant populates the intermediate state, 1-anilino-8-napthalene sulphonic acid (ANS) binding was monitored by fluorescence (Figure 1c). The RNase H kinetic intermediate resembles its equilibrium molten globule[6], and one of the trademarks of molten globules is binding of the hydrophobic dye ANS. When ANS binds exposed hydrophobic surface area, its emission maximum blue-shifts and its fluorescence intensity increases[23]. Under native conditions, I25A incubated with ANS shows a large increase in fluorescence compared to the wild type enzyme; this increase is similar in magnitude to what is generally observed for molten globules.

I25A RNase H populates the native fold

In contrast to the evidence presented above, several studies demonstrate that I25A retains a native-like fold. The CD spectrum of I25A under native conditions is indicative of a folded, α/β protein and resembles that of wild type RNase H (Figure 2a). I25A RNase H is also active, assessed by the loss of the hypochromic effect as the RNA strand is cleaved from a DNA-RNA hybrid (Figure 2b). Preservation of the active enzyme, albeit impaired in comparison to wild type, implies that the protein can access the native fold under these conditions. I25 is completely buried in the periphery and makes no contact with substrate.

To obtain higher resolution structural information, a native state ^{15}N-^{1}H HSQC NMR spectrum was obtained (Figure 2c). I25A shows clearly dispersed peaks characteristic of a well-folded protein; in contrast, the spectra of molten globules and unfolded proteins exhibit a collapsed set of peaks in the ^{1}H dimension. Furthermore, the I25A peaks overlay well with those of wild type RNase H, clearly demonstrating that I25A populates the native structure under the conditions of the NMR experiments.

The two-state assumption is not valid for I25A

In light of the conflicting behavior of I25A RNase H, intrinsic tryptophan fluorescence was used to further probe equilibrium behavior via urea denaturation. RNase H has six Trp residues. Fluorescence intensity at 340 nm as a function of urea concentration was fit using the two-state approximation allowing a direct comparison between CD and fluorescence. In the case of wild type RNase H, the curves representing the fraction folded as calculated by CD and fluorescence are indistinguishable[20]. For I25A, however, the fluorescence and CD data are non–coincident (Figure 3a). The fit from the fluorescence results in an m-value of 1.5 kcal mol^{-1} M^{-1} and ΔG_{UN} of 4.1 kcal/mol, which, while again significantly lower than wild type values, do not match with those calculated by CD. This non-coincidence of probes suggests that for I25A RNase H the two-state assumption breaks down. In support of this, the emission spectra do not have an isosbestic point, also indicating that more than two species contribute to the signal.

To obtain an estimate of the true global stability of the native conformation without assuming a two-state model, hydrogen exchange was carried out on the I25A variant. By monitoring the rate at which amide hydrogens exchange with solvent deuterium by NMR, we calculated the free energy of exchange for 63 residues ($\Delta G = -RT \ln (k_{obs}/k_{rc})$)[24]. The ΔG_{HX} values obtained for I25A, shown in Figure 3b, reveal that the regions with the highest protection are the A and D helices, located in the core. These values suggest a global stability of approximately 6 kcal/mol. Since it is known that D_2O has an effect on protein stability[25], we repeated the denaturant melt in D_2O in order to compare the stability calculated by hydrogen exchange to that obtained by standard urea denaturation (data not shown). The stability of I25A in D_2O, assuming a two-state model, is increased by less than 1 kcal/mol, and the m-value remains the same. A large discrepancy, therefore, remains between the stability of I25A calculated assuming a two-state fit and that by hydrogen exchange (Figure 3b). This incongruity is not evident for wild type RNase H. In this case the values for the global stability of the protein determined by native state hydrogen exchange and a two-state fit to the equilibrium denaturation profile are closely matched[9], confirming the validity of the two-state assumption. The inconsistency observed for I25A is additional evidence that the two-state assumption does not hold for this variant.

I25A populates the intermediate under native conditions

We determined the effect of I25A in the background of D10A, a mutation in the periphery known to stabilize RNase H by 3.3 kcal/mol (ΔG_{UN} = 13.0 kcal/mol)[5]. In a more stable background, there is less potential for a destabilizing mutation to affect the apparent two-state equilibrium transition. Equilibrium urea denaturation of D10A/I25A was monitored by CD and fit using the two-state approximation (Figure 4a). The introduction of I25A into D10A destabilizes the protein by 5.1 kcal/mol (ΔG_{UN} = 7.9 kcal/mol) and results in an m-value of 2.1 kcal mol^{-1} M^{-1}, identical to that of wild type

and D10A (Table 1). Assuming additivity provides a theoretical value for the stability of I25A of 4.6 kcal/mol (Scheme 1).

We constructed a model of the populations of native, intermediate, and denatured I25A RNase H as a function of urea using this value for ΔG_{UN} and the wild type values for m_{UN}, ΔG_{UI} and m_{UI}. This model, shown in Figure 4b, reveals that the stability of the native state of I25A is sufficiently close to the stability of the intermediate that a significant population of intermediate exists under native conditions and becomes the dominant species around 2 M urea.

NMR spectra in denaturant reveal two sets of peaks.

Confident that I25A does in fact populate the intermediate state, we obtained HSQCs at varying urea concentrations in the hopes of obtaining structural information about the intermediate at the residue level (Figure 5a-e). In low urea, the spectrum closely resembles that in no denaturant. Upon closer inspection, especially around the glycine region, however, there is a clear cluster of peaks that is not present in the wild type spectrum under native conditions. These peaks are circled in red in Figure 2c and Figure 5a-3. As urea is added, these peaks intensify and persist to high concentrations of denaturant, indicating that they correspond to residues in the unfolded state.

Peaks in other regions of the spectrum also emerge at increasing urea concentrations. By 1.5 M urea (Figure 5c), there are clearly two sets of peaks. One set corresponds to the dispersed set expected for the native conformation, and these decrease in intensity as denaturant increases. A second set, which builds in intensity with urea concentration, is characterized by a lack of dispersion in the ^{1}H dimension and appears to arise from residues in the unfolded state. This is in stark contrast to what we observe for the wild type protein, where we see one set of peaks at all concentrations of urea.

3.4 DISCUSSION

Protein engineering when combined with prior knowledge of the energy landscape of a protein, particularly from hydrogen exchange data, can be a useful tool to manipulate the population of protein conformations. Here, by creating a single destabilizing mutation in the periphery of RNase H (I25A) we show that it can be used to the protein populates the partially folded intermediate known to be important in the folding trajectory of the protein, and thereby allowing us to study this experimentally elusive species at equilibrium.

Modeling the population of each species in I25A

Initial characterization of I25A provided conflicting views of the effect of the mutation. The CD spectrum and ^{15}N-^{1}H HSQC indicate that the overall structure of the protein is not notably perturbed from that of wild type. The activity of I25A, although decreased compared to wild type, also indicates that the native fold is accessed. On the other hand, the significantly lower m-value obtained by the two-state fit of the denaturant melts and the ANS binding assay suggested that I25A resembles the intermediate conformation of RNase H. The non-coincidence of CD and fluorescence and the estimate of global stability obtained by hydrogen exchange indicate that the two-state approximation cannot be applied in this case. Examining the effect of the isoleucine to alanine mutation at position 25 in a more stable background (D10A) allowed us to obtain a better estimate of the effect of this mutation on the native state of RNase H. Surprisingly, this single mutation appears to destabilize the protein by more than 5

kcal/mol. Since the stability of the intermediate of RNase H is approximately 3.5 kcal/mol, a mutation of this magnitude that selectively destabilizes the native state (~10 kcal/mol) will result in a native state stability close to that of the intermediate and in a significant accumulation of the intermediate at equilibrium. The two-state assumption used to fit the denaturant melt is therefore invalid in this case, resulting in an inaccurate determination of the stability of the protein standard urea-induced denaturation profiles.

Why does this single amino acid change have such a drastic effect on the stability of RNase H? By introducing alanine at position 25, we expected to destabilize the periphery by creating a small cavity in the hydrophobic pocket. Based on the change in hydrophobic surface area, we expect this mutation to destabilize the protein by ~ 3 kcal/mol[26]. At this point, we do not fully understand why the change is so much greater, although such large changes are not without precedent. For example, in T4 lysozyme, small perturbations (leucine to alanine) at buried sites have been known to result in large stability changes when coupled to the formation of a cavity[27]. In our case, this effect may also be exacerbated by the presence of a nearby salt bridge network. Creating a cavity in the vicinity of these salt bridges may decrease the strength of the interaction, allowing the periphery to unfold more readily.

Our three-state equilibrium model successfully accounts for all of our experimental data. To model our data, the intermediate was assigned a CD signal 80% that of the native state based on kinetic studies on the wild type protein[6]. Fitting the simulated melt to the two-state model reproduces the observed data with a calculated free energy of unfolding (3.5 kcal/mol) and the m-value (1.3 kcal mol^{-1} M^{-1}) of I25A RNase H.

The stability estimate obtained by hydrogen exchange is also consistent with our conclusions. This method should allow us to measure the free energy of hydrogen exchange of individual residues with an upper limit of the stability of I25A in D_2O. By looking at the most protected residues in the protein, those in helices A and D, we would estimate the stability of the protein to be ~ 5 - 6 kcal/mol. Studies of RNase H in the presence of D_2O suggest that the solvent increases the stability by just over 1 kcal/mol, which would put our estimate of the stability of I25A at ~4.5 kcal/mol, consistent with the value derived from assuming additivity with the mutation D10A. We attempted to repeat the urea melts used in this study in D_2O to get a more exact number, but urea is not stable for the long equilibration times required for the variants containing D10A. Nonetheless, our expectation of 4.5 kcal/mol is consistent with both the hydrogen exchange and mutant cycle analysis.

The HSQC spectra in varying urea concentrations provide further corroborating evidence that I25A populates the intermediate. The cluster of peaks circled in Figure 2c is not present in the wild type spectrum. They fall in the glycine region, and twelve of the fifteen glycines in RNase H are located in the periphery, a region presumed to be unstructured in the intermediate. These peaks persist in the higher urea concentrations (see peaks circled in red, Figures 2c and 5), where two sets of peaks are evident, one corresponding to folded protein and one characterized by the lack of dispersion typical of unfolded proteins. At these relatively low levels of denaturant, there should be no detectable unfolded protein, suggesting that these peaks arise from the intermediate state. In addition, similar spectra of the wild type protein in varying amounts of urea do not

exhibit two sets of peaks. We therefore attribute the glycine peaks and the appearance of the collapsed second set to unfolded regions of the intermediate.

No peaks appear to correspond to structured areas of the intermediate. The dispersed set of peaks appears to arise from the native state. At 2 M urea, where the intermediate should be maximally populated, the dispersed set of peaks overlays well with the native spectrum, and disappear together with the glycine peaks believed to arise from the native periphery. Together, these data suggest that the peaks from the folded core arise from the native structure, not the folded region of the intermediate. Although this set of peaks does exhibit some shift from the native spectrum, these shifts may arise from the addition of urea and are not of the magnitude we would expect if they were to originate from the intermediate. The gradual shift of peaks can also indicate two species in fast exchange, the observed peak being an average of the separate signals from each of the populations rapidly interconverting. The intermediate and native structure are unlikely to be in fast exchange because the rate of the I → N step in H_2O is 0.74 s^{-1}, far slower than the NMR fast exchange regime. Instead we attribute the loss of peaks for the structured region of the intermediate to dynamic line broadening.

Insights into the nature of the folding intermediate

The HSQC spectra of molten globules commonly show no peaks that report on the structured regions, with signal building in only as structure is lost[28]. Fluctuations throughout the molten globule on the millisecond timescale are cited as the cause of peak broadening. We believe that, likewise, the absence of peaks arising from the intermediate in these experiments indicates that it is a dynamic species. This exchange between conformers suggests that the barriers between them are low and that they lack fixed tertiary interactions; contacts in the core, though native-like, are most likely weak.

This model of the folding intermediate of RNase H as a dynamic species agrees with studies on equilibrium molten globules that are thought to mirror the intermediates formed in the kinetic folding trajectory. The intermediate of apomyoglobin is also thought to be a heterogeneous and dynamic ensemble[29]. Our results, however, differ from those obtained on other systems where mutations were made to selectively populate the intermediate, such as Im7 and redesigned-apocytochrome b_{562}. Both of these proteins are four-helix proteins with one or more helices unstructured in their intermediates. NMR studies on the models of these intermediates indicate that the folded regions are quite rigid[18;30;14;31], although the Im7 intermediate does show increased heterogeneity over the native state. In the cases of both Im7 and redesigned-apocytochrome b_{562}, the authors stress the existence of non-native contacts in the intermediate that need to be broken upon folding to the native state. Instead of utilizing non-native tertiary contacts to stabilize the intermediate, RNase H appears to fluctuate, forming weak native-like interactions that do not persist.

Our results are also at odds with recent NMR studies on a fragment of *T. thermophilus* RNase H that forms a well-folded native-like subdomain. The fragment was generated by removing regions from the periphery: two internal strands and the final helix of the protein, both of which appear unprotected by hydrogen exchange, while leaving one strand from the periphery (strand 1) which, based on native state hydrogen exchange and mutagenesis studies, also does not appear to be structured in the folding intermediate of *E. coli* RNase H. The inclusion of this strand may stabilize the fragment, allowing for a well-behaved sample amenable to high resolution NMR. The high-

resolution structure reveals only native-like interactions, leading to questions about the nature of the barrier in the progression of folding from the intermediate to the native state and at odds with the observation that mutations in this region affect the transition state stability more than the intermediate[5,7]. The difference between what we find here and the NMR study on this fragment may also be attributed to a difference in the interactions that stabilize the intermediates from these two proteins (*E. coli* and *T. thermophilus*) . Another explanation might be that the well-folded fragment is actually a mimic of another identified high-energy intermediate of the *T. thermophilus* protein where just the last helix appears to be unfolded.

The dynamic nature of the *E. coli* RNase H intermediate may have important implications for efficient and productive folding in vivo. The intermediate of RNase H is populated prior to the rate-limiting step and its presence appears to aid in folding. Destabilizing the intermediate slows overall folding, suggesting that the interactions that stabilize the native state are present and important in the rate-limiting transition state. If interactions that form as the reaction progresses stabilize the transition state more than the intermediate, the folding reaction will be accelerated. This may be the case for RNase H, for which all evidence points to a productive, on-pathway, obligatory intermediate.

3.5 MATERIALS AND METHODS

Materials

Deuterium oxide, ^{15}N ammonium chloride, deuterated buffers, acids, and bases were purchased from Isotec. All other buffer reagents were purchased from Sigma or Fischer.

Protein Expression and Purification

E. coli BL21 pLysS cells were transformed with the appropriate plasmid and grown at 37°C in LB medium with 200 μg/ml ampicillin. Induction was initiated by the addition of 1 mM IPTG to cells at OD_{600} ~0.6. Cells were harvested by centrifugation 3 – 4 hours after induction. I25A expressed in inclusion bodies, and purification was carried out on cell pellets as previously described for other RNase H variants[7]. D10A/I25A was expressed solubly and was purified as previously described[20]. Purity and molecular weights of all variants were verified by mass spectrometry (data not shown).

To express ^{15}N-labeled protein, log-phase cells were transferred to M9 medium containing ^{15}N ammonium chloride as described[32].

All experiments were carried out in 20 mM sodium acetate and 50 mM KCl at pH 5.5 unless otherwise noted. Protein concentrations were determined based on the extinction coefficient, calculated according to the number of Trp and Tyr residues[33].

Equilibrium CD Experiments

Circular dichroism measurements were carried out on an Aviv 62DS spectrometer at 25°C. For denaturant melts, individual samples containing 40 - 50 μg/mL protein at varying urea concentrations were equilibrated overnight. For each sample, CD signal was monitored at 222 nm, and the signal was averaged over a 60 second time period. Data were fit assuming a two-state model and linear dependence of ΔG_{UN} on urea concentration.

ANS Binding

Samples containing 500 μM 1-anilino-8-napthalene sulphonic acid in buffer with and without 2 μM protein were prepared and equilibrated overnight. Fluorescence emission spectra were collected from 425 to 675 nm with an excitation wavelength of 405 nm. The spectrum of ANS in buffer alone was subtracted from those of ANS containing protein.

Activity Assay

Activity of RNase H is monitored by the loss of the hypochromic effect as the RNA strand is cleaved from a DNA-RNA hybrid. The reaction was carried out in 50 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, and 10 μg/mL rA-dT and was initiated by the addition of enzyme to a final concentration of 5 nM. Absorbance at 260 nm was followed over time on a Cary UV-Vis spectrometer.

Tryptophan Fluorescence Measurements

Urea denaturation was monitored by tryptophan fluorescence using a Fluoromax 3 fluorimeter (JYHoriba) at 25°C. Individual samples of 40 – 50 μg/mL in varying urea concentrations equilibrated overnight. Excitation was at 295 nm, and emission spectra were recorded with both slits at 4 nm. Fluorescence at 340 nm as well as the center of mass were analyzed and fit using a two-state approximation and a linear dependence of ΔG_{UN} on urea concentration.

Hydrogen Exchange

Amide hydrogen exchange was initiated by exchanging protonated ^{15}N I25A RNase H into deuterated buffer (20 mM sodium acetate and 50 mM KCl at pDr 5.6) using a polypropylene spin column (Pierce) packed with Sephadex resin. The sample was immediately transferred to an NMR tube and placed in the instrument; time between initiation of exchange and start of data collection was approximately 25 minutes. ^{15}N-^{1}H HSQC spectra were recorded on a Bruker 600 MHz at 25°C as an average of 16 scans with 1024 points in the direct dimension and 256 complex points in the indirect dimension. HSQC's (~1 hour each) were collected consecutively for 10 hours and then increasingly spaced out to two weeks. Spectra were processed using Felix 97.0 (Accrelys), and peak height as a function of time was fit to a single exponential decay in SigmaPlot (SSI) to obtain a value for k_{obs}. Wild type peak assignments were used to assign peaks for I25A. The free energy of exchange was calculated as:

$$\Delta G = -RT \ln (k_{obs}/k_{rc})$$

where k_{rc} is the intrinsic rate of exchange for that residue in a random coil [24].

HSQC's at varying concentrations of urea

Two-dimensional, gradient-enhanced HSQC's were recorded on a Bruker 600 MHz at 25°C. 32 scans were collected with 1024 points in the direct dimension and 128 complex points in the indirect dimension. These data were processed using NMRPipe and viewed in NMRDraw.

3.6 ACKNOWLEDGEMENTS
We thank David E. Wildes for suggestions and development of this project and Rachel Bernstein for thoughtful comments on the manuscript. We also acknowledge David Wemmer for NMR use and Jeff Pelton for guidance in NMR data collection. This work was supported by NIH grant GM50945 to S.M. and an NSF graduate research fellowship to K.B.C.

3.7 REFERENCES

1. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60.
2. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**, 321-31.
3. Bose, H. S., Whittal, R. M., Baldwin, M. A. & Miller, W. L. (1999). The active form of the steroidogenic acute regulatory protein, StAR, appears to be a molten globule. *Proc Natl Acad Sci U S A* **96**, 7250-5.
4. Jahn, T. R., Parker, M. J., Homans, S. W. & Radford, S. E. (2006). Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol* **13**, 195-201.
5. Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* **6**, 825-31.
6. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* **4**, 298-304.
7. Spudich, G. M., Miller, E. J. & Marqusee, S. (2004). Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol* **335**, 609-18.
8. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057-60.
9. Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* **3**, 782-7.
10. Spudich, G., Lorenz, S. & Marqusee, S. (2002). Propagation of a single destabilizing mutation throughout the Escherichia coli ribonuclease HI native state. *Protein Sci* **11**, 522-8.
11. Goedken, E. R. & Marqusee, S. (2001). Native-state energetics of a thermostabilized variant of ribonuclease HI. *J Mol Biol* **314**, 863-71.
12. Bai, Y. (2006). Energy barriers, cooperativity, and hidden intermediates in the folding of small proteins. *Biochem Biophys Res Commun* **340**, 976-83.
13. Bai, Y., Feng, H. & Zhou, Z. (2007). Population and structure determination of hidden folding intermediates by native-state hydrogen exchange-directed protein engineering and nuclear magnetic resonance. *Methods Mol Biol* **350**, 69-81.
14. Feng, H., Vu, N. D. & Bai, Y. (2004). Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome b562: native fold with non-native hydrophobic interactions. *J Mol Biol* **343**, 1477-85.
15. Feng, H., Vu, N. D. & Bai, Y. (2005). Detection of a hidden folding intermediate of the third domain of PDZ. *J Mol Biol* **346**, 345-53.
16. Kato, H., Feng, H. & Bai, Y. (2007). The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J Mol Biol* **365**, 870-80.

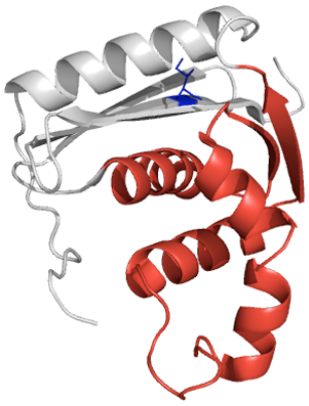
17. Kato, H., Vu, N. D., Feng, H., Zhou, Z. & Bai, Y. (2007). The folding pathway of T4 lysozyme: an on-pathway hidden folding intermediate. *J Mol Biol* **365**, 881-91.
18. Spence, G. R., Capaldi, A. P. & Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**, 215-26.
19. Chamberlain, A. K. & Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry* **37**, 1736-42.
20. Dabora, J. M. & Marqusee, S. (1994). Equilibrium unfolding of Escherichia coli ribonuclease H: characterization of a partially folded state. *Protein Sci* **3**, 1401-8.
21. Greene, R. F., Jr. & Pace, C. N. (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* **249**, 5388-93.
22. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138-48.
23. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119-28.
24. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86.
25. Makhatadze, G. I., Clore, G. M. & Gronenborn, A. M. (1995). Solvent isotope effect and protein stability. *Nat Struct Biol* **2**, 852-5.
26. Xu, J., Baase, W. A., Baldwin, E. & Matthews, B. W. (1998). The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci* **7**, 158-77.
27. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178-83.
28. Redfield, C. (2004). Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34**, 121-32.
29. Nishimura, C., Dyson, H. J. & Wright, P. E. (2002). The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate. *J Mol Biol* **322**, 483-9.
30. Gsponer, J., Hopearuoho, H., Whittaker, S. B., Spence, G. R., Moore, G. R., Paci, E., Radford, S. E. & Vendruscolo, M. (2006). Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc Natl Acad Sci U S A* **103**, 99-104.
31. Feng, H., Zhou, Z. & Bai, Y. (2005). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc Natl Acad Sci U S A* **102**, 5026-31.
32. Marley, J., Lu, M. & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-5.
33. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948-54.

Table 3.1: Stability and m-values for RNase H variants. The values in parentheses represent the standard deviation calculated from at least three individual denaturant melts. The reported ΔG_{UN} for D10A/I25A was obtained by fixing the m-value to 2.1 kcal mol^{-1} M^{-1}. See accompanying manuscript for details.

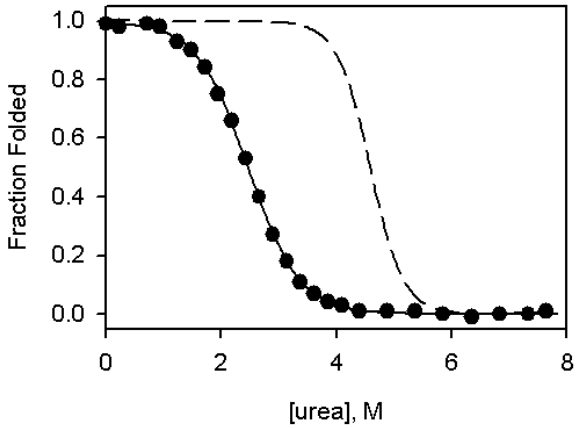
	wt[6]	I53A[5]	D10A[5]	I25A		D10A/I25A
				urea melt	model	
ΔG_{UN} (kcal mol^{-1})	9.7	7.6	13	3.4 (0.3)	4.5	7.86 (0.07)
m_{UN} (kcal mol^{-1} M^{-1})	2.1	2.1	2.1	1.3 (0.1)	2.1*	2.1
ΔG_{UI} (kcal mol^{-1})	3.55	1.66	4.38	ND	3.55*	ND
m_{UI} (kcal mol^{-1} M^{-1})	1.24	1.24	1.45	ND	1.24*	ND

*Wild type values assumed for the model
ND: not determined

(a)



(b)


Fraction Folded
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
[urea], M

(c)

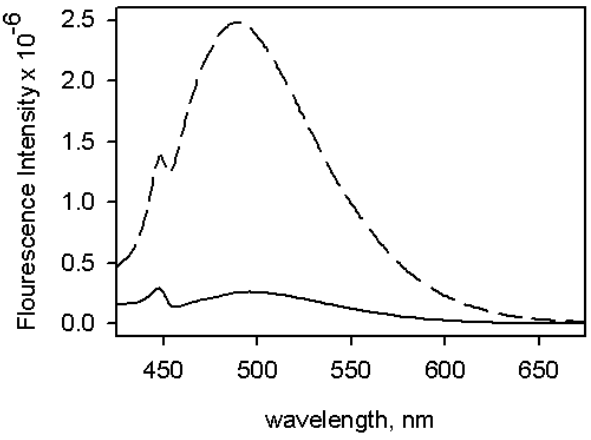
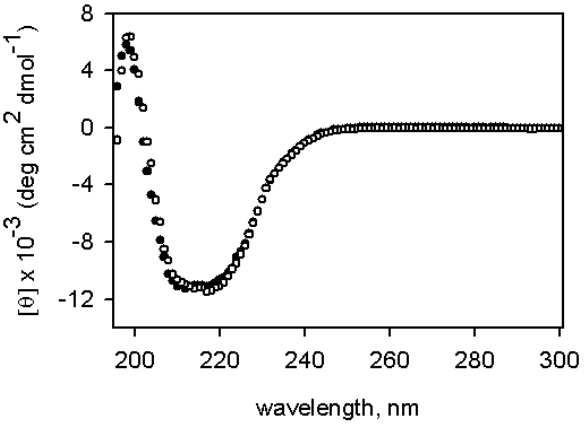
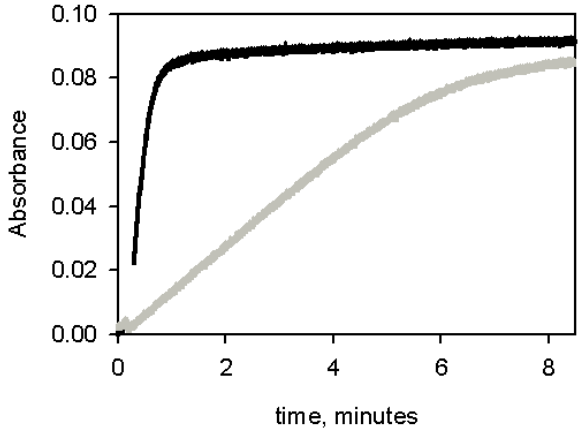

Flourescence Intensity x 10-6
2.5
2.0
1.5
1.0
0.5
0.0
450
500
550
600
650
wavelength, nm

Figure 3.1: I25A resembles the RNase H intermediate. (a) Ribbon representation of RNase H (pdb 1F21) with the core of the protein colored red and the periphery colored grey. Residue I25 is shown as sticks and colored blue. (b) Fraction folded of I25A (filled circles) determined from urea melts overlaid with the wild type curve (dashed line). The solid line represents the two-state fit for I25A. (c) Fluorescence emission spectra of ANS in the presence of WT (solid line) and I25A (dashed line). The fluorescence emission of ANS in buffer alone has been subtracted from the data.

(a)


8
4
0
-4
-8
-12
200
220
240
260
280
300
wavelength, nm
[θ] x 10^-3 (deg cm^2 dmol^-1)

(b)


0.10
0.08
0.06
0.04
0.02
0.00
0
2
4
6
8
time, minutes
Absorbance

(c)

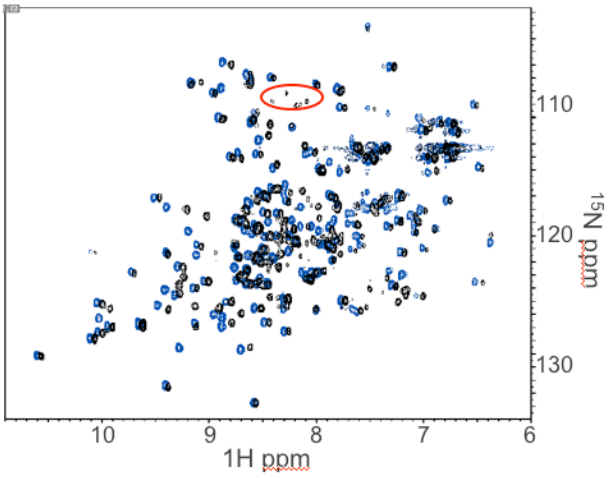
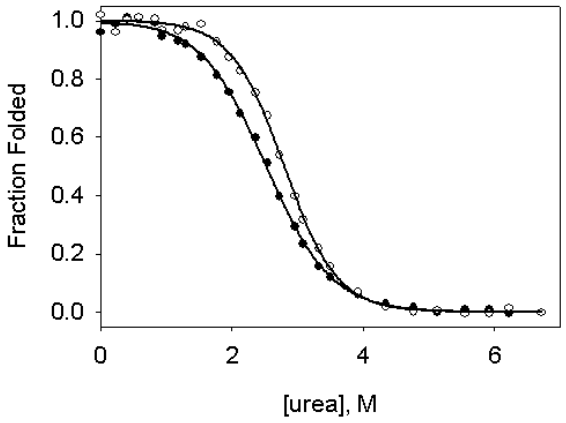

110
120
130
15N ppm
10
9
8
7
6
1H ppm

Figure 3.2: Evidence for an E. coli RNase H fold. (a) CD spectrum of I25A (filled circles) overlaid with wild type RNase H (open circles). (b) Activity of I25A (grey) compared with that of WT (black). (c) ^{1}H-^{15}N HSQC of I25A RNase H (black) overlaid with that of WT (blue) in 20 mM NaOAc, pH 5.5, 50 mM KCl, 10 % D_2O.

(a)



(b)

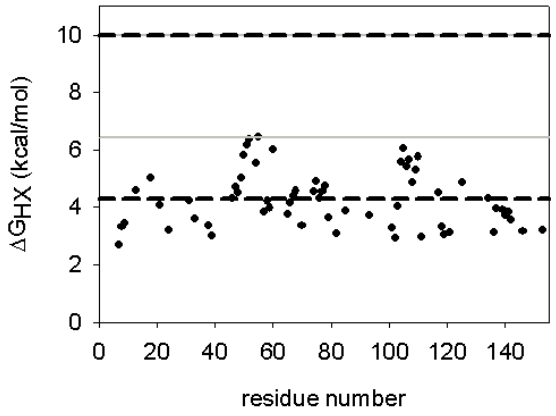
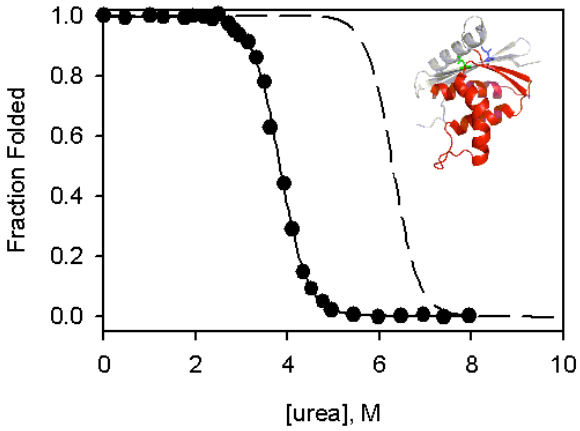
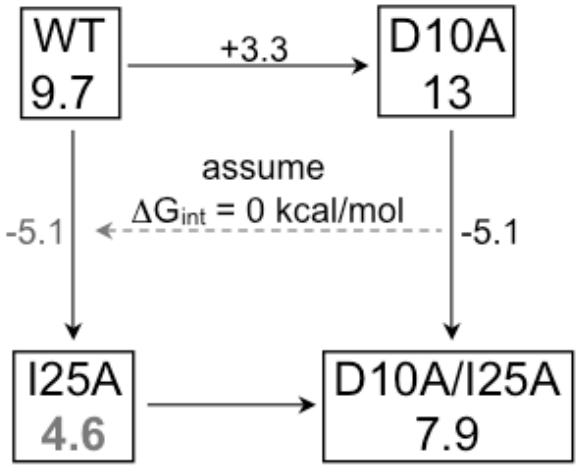


Figure 3.3: I25A does not follow two-state behavior at equilibrium. (a) Fraction folded of I25A by CD (closed circles) and fluorescence (open circles) determined from urea melts fit to a two-state approximation. Solid lines represent each fit. (b) Hydrogen exchange data for I25A. The free energy of exchange of I25A is plotted as a function of residue. The grey solid lines represent the stability of WT and I25A by hydrogen exchange methods, as indicated, and the dashed black lines represent the stability of each protein calculated from the two-state fits to melts acquired in deuterated solvent.

(a)


1.0
0.8
0.6
0.4
0.2
0.0
Fraction Folded
0
2
4
6
8
10
[urea], M

(b)


WT
9.7
+3.3
D10A
13
assume
ΔGint = 0 kcal/mol
-5.1
-5.1
I25A
4.6
D10A/I25A
7.9

(c)

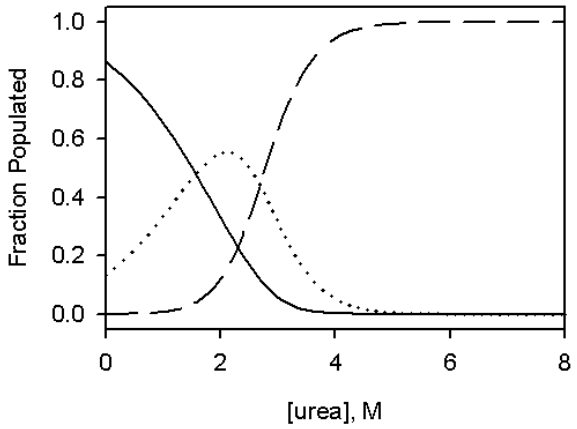


Figure 3.4: Determining the effect of I25A. (a) Fraction folded of D10A/I25A determined from urea melts (filled circles) overlaid with curves for wild type (dashed line). The solid line represents the two-state fit to the D10A/I25A melt. Inset: Locations of I25 (blue) and D10 (green) shown on the structure of RNase H with the periphery in grey and the core in red. (b) A thermodynamic cycle assuming no interaction between residues 10 and 25 was used to calculate the expected stability of I25A. Numbers in each box represent the global stability, ΔG_{UN} (kcal/mol), and the numbers along the arrows indicate the change in stability upon making the indicated mutation. (c) A graph of the theoretical populations for the native (solid), intermediate (dotted), and unfolded (dashed) forms of I25A as a function of urea concentration. The model assumes the native form of I25A is 4.6 kcal mol^{-1} stable with respect to the denatured form and that an m-value of 2.1 kcal mol^{-1} M^{-1} is associated with this transition. The intermediate is assumed to be 3.5 kcal mol^{-1} stable with respect to the denatured form, as it is in the wild type protein, with an m-value of 1.3 kcal mol^{-1} M^{-1}.

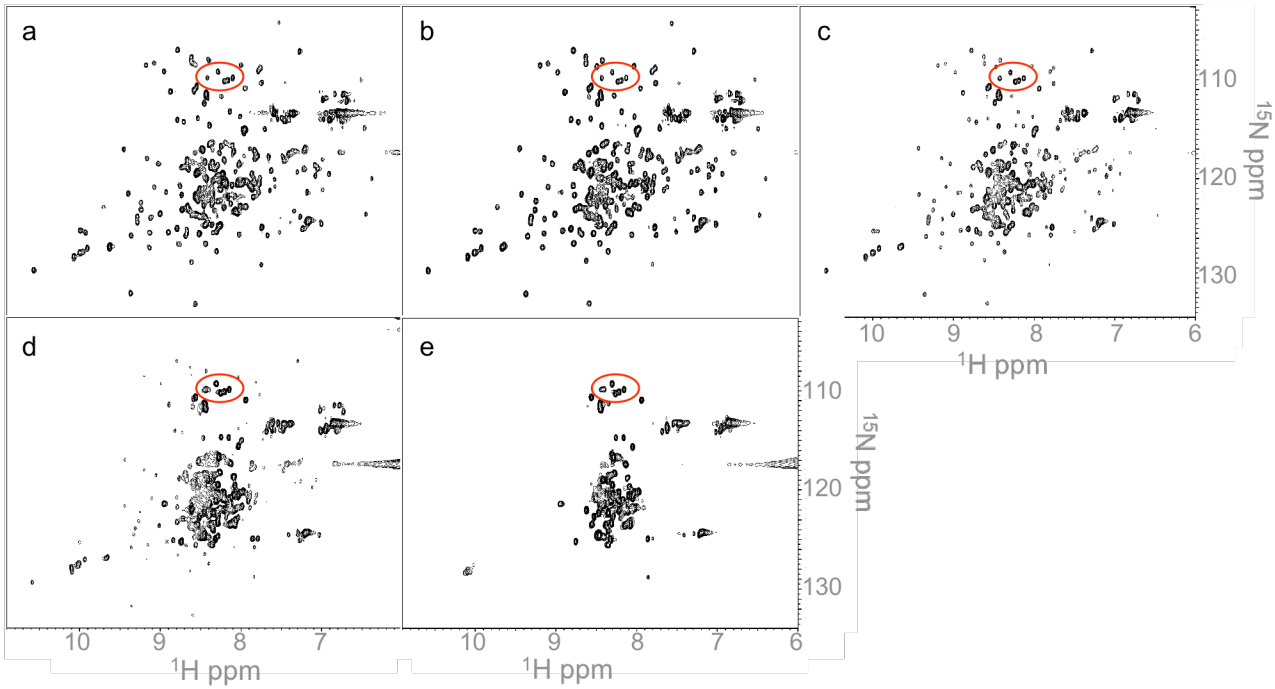


Figure 3.5: Probing the intermediate by NMR. A series of HSQC's in 20 mM NaOAc, pH 5.5, 50 mM KCl, 10 % D_2O and (a) 0.5 M, (b) 1.0 M, (c) 1.5 M, (d) 2.0 M, and (e) 2.5 M urea.

Chapter IV

Equilibrium models of kinetic folding intermediates: A comparison of the intermediates of *E. coli* and *T. thermophilus* RNases H

4.1 ABSTRACT

Protein folding intermediates, usually inaccessible to direct structural characterization, have long been a source of debate. High-resolution structural information would provide invaluable information regarding the role and nature of these controversial species. *E. coli* ribonuclease H is known to fold through a productive and obligatory intermediate. Here, we attempt to populate this intermediate at equilibrium through multiple destabilizing mutations in the regions thought to be uninvolved in the intermediate. This stability-biasing approach was applied to both *E. coli* and *T. thermophilus* RNases H in order to compare properties of the high-energy forms of homologous proteins that function in vastly different environments. Equilibrium CD and NMR experiments indicate that we have successfully constructed mimics of the partially folded intermediates of these proteins. Preliminary NMR data on these intermediate mimics pose many interesting questions to pursue.

4.2 INTRODUCTION

Partially folded intermediates have long been recognized as species important to the protein folding process. The role of such intermediates, however, remains widely debated[1; 2]. Do they solely represent off-pathway species that lead to aggregation and disease, or do they promote folding by narrowing the conformational search[2; 3]? Characterization of kinetic intermediates by hydrogen exchange[4; 5], phi-value analysis[6], and other methods has been enlightening, providing a wealth of information about the role of topology[7; 8] and local stability[9; 10] in defining the folding trajectory. Recent improvements in fast mixing and NMR techniques have resulted in the identification of intermediates on the pathways of proteins previously thought to be two-state folders, suggesting that these high-energy species are a general feature of the energy landscapes of even small proteins[11; 12; 13].

Studies of the folding process of homologous proteins suggest that such intermediate ensembles are likely to be ubiquitous during the folding process. For instance, the immunity protein Im7 populates a partially folded intermediate, while Im9, sharing 60 % sequence identity, does not[14]. A change in conditions or rational design of mutants, however, results in the observation of a similar kinetic intermediate during the folding of Im9[15]. A switch in apparent mechanism by changing a few residues or conditions has been observed for several other proteins as well[16; 17; 18; 19]. These studies suggest that energy landscapes are inherently rough, and identifying intermediates can simply be a matter of making the features on a landscape detectable by standard methods. Even when they are detectable, however, intermediates are weakly and transiently populated, prohibiting a detailed structural analysis.

Here, by creating destabilizing mutations that selectively destabilize the native state and thereby populate the intermediate, we constructed equilibrium mimics of the kinetic intermediates of RNases H from *E. coli* and *T. thermophilus*. In addition to providing insight into the mechanism of protein folding, a comparison of the intermediates in proteins from organisms that live in distinct environments can lend considerable insight into how relatively few sequence differences confer important differences on the energy landscape. RNases H from *E. coli* and *T. thermophilus* share 52 % sequence identity and indistinguishable structures (Figure 1)[20; 21]. Both fold through kinetic intermediates composed of a structured core region and disordered

periphery, as characterized by quenched-flow hydrogen exchange[10; 22] and phi-value analysis[23]. The kinetic intermediate of *E. coli* RNase H is known to play a robust and important role in folding, dominating even the trajectory of variants that show no apparent accumulation of the species[24]. Single molecule experiments have shown this intermediate to be on-pathway and obligatory[25].

In spite of sharing many important similarities, RNase H from *T. thermophilus* must remain folded and function at temperatures at which the *E. coli* protein would be mostly unfolded[26]. Residual structure in the denatured state of *T. thermophilus* RNase H, suggested by its lower ΔC_p, has been implicated in its mechanism of thermostabilization[27; 28]. Recent results from our lab and the Bai lab led us to ask if the dynamic properties of their kinetic intermediates might be another important difference between the mesophilic and thermophilic RNases H. The two groups used different approaches to create equilibrium mimics of the kinetic intermediates. We set out to selectively destabilize the native state of *E. coli* RNase H by introducing mutations only into the periphery, the region believed to be unfolded in the intermediate, thereby enhancing the population of the intermediate[29]. Surprisingly, a single destabilizing mutation, I25A, was sufficient to accomplish this goal. Characterization of this mutant by ANS binding and a series of HSQC's suggested that the intermediate is a broad and dynamic ensemble of conformations exchanging on the millisecond timescale. Alternatively, Bai and coworkers generated a fragment of *T. thermophilus* RNase H to represent the partially folded form by removing the regions thought to be disordered in this species[30]. The result was a rigid, well-packed protein, whose NMR structure revealed only native-like interactions.

The contradicting views of the dynamic properties of the intermediates of *E. coli* and *T. thermophilus* RNases H left us with many questions, the first stemming from the inclusion of a strand from the periphery (Strand I) in the fragment of *T. thermophilus* RNase H. Native state and quenched flow hydrogen exchange experiments, on which the fragment was based, do not provide a clear structural role for Strand I in the intermediate[22; 31]. In the quenched flow hydrogen exchange studies, partial protection was observed at the earliest refolding time for the single probe in Strand I. When interpreting the native state hydrogen exchange experiments, Hollien et al. declined to place Strand I in a specific PUF; one of its three measurable residues had a ΔG_{unf} very similar to other probes in the periphery while two were closer to the core. Is a fragment of *T. thermophilus* RNase H that includes Strand I an accurate representation of the kinetic intermediate, and is the inclusion of Strand 1 responsible for the apparent difference in packing between the two mimics? The observed difference may also arise from the difference in approach, fragmentation versus stability-biasing. The 'disordered' periphery may affect the dynamics of the folded core region. It is also possible, of course, that the intermediates of these two proteins behave very differently; if so, how is this important to the overall energy landscapes and ability to function at high temperatures?

In attempt to address these questions and perhaps reconcile the two studies, we set out to create equilibrium mimics of *E. coli* and *T. thermophilus* RNases H using the stability-biasing approach. Since I25A *E. coli* RNase H forms a mixed population of the intermediate and native forms, we sought to further populate the intermediate by adding mutations to selectively destabilize N. To investigate the role of Strand I in these two intermediates, two versions of the intermediate were created for each protein: one that

does and one that does not include mutations in Strand I. Here we report the successful construction of mimics of kinetic intermediates and preliminary structural information obtained by NMR.

4.3 RESULTS

Construction of intermediate mimics

To create partially unfolded proteins that represent the kinetic intermediates of *E. coli* and *T. thermophilus* RNases H, periphery residues that do not contact the core were targeted for mutation. Single-site mutants were made and purified, and the destabilizing effect of each mutation was assessed by urea denaturation[32]. The ΔΔG's for each single-site variant, summarized in Table 1 for *E. coli* RNase H and Table 2 for *T. thermophilus* RNase H, were then treated additively to estimate the native state stability of the protein that would result from combining the given mutations. To preserve additivity to our best ability, residues contacting each other in the crystal structures of the proteins were not mutated simultaneously in a given intermediate mimic. For instance, when the strand I mutation F8A was introduced in each protein, the identity of residue 25 was kept as wild type, as these two residues occupy the same hydrophobic pocket in the crystal structure. When possible, mutations at analogous sites in *E. coli* and *T. thermophilus* RNases H were used. The theoretical stability of the "native" state of each of the mimics was designed to be close to that of the unfolded state in the hopes of obtaining a variant with the intermediate as the ground state and to avoid a mixture of populations.

Two versions of intermediate mimics were created for each RNase H: one does not contain any mutations in Strand I and one does have mutations in Strand I. These will be referred to herein as EFLwt1 (for *E. coli* RNase H Full-Length intermediate mimic with wild type Strand I), EFLmut1 (for *E. coli* RNase H Full-Length intermediate mimic with destabilizing mutations in Strand I) and analogously for *T. thermophilus* RNase H, TFLwt1 and TFLmut1. The point mutations used in this study and their effects on the stability of the native state are summarized in Tables 1 and 2.

Equilibrium CD experiments

All four mimics showed CD spectra indicating formation of secondary structure (Figure 2). The CD spectra of these mimics, however, do not resemble folded RNase H, suggesting that there are gross structural perturbations relative to wild type.

Equilibrium urea denaturation was carried out for all four intermediate mimics, and the resulting curves were fit using linear extrapolation and the two-state assumption to determine their stabilities assuming a two-state model (Figure 3)[32]. Both *E. coli* RNase H intermediate mimics have similar stabilities and m-values (3.3 kcal mol^{-1} and 1.2 kcal mol^{-1} M^{-1}, respectively). These values match those calculated for the wild-type kinetic intermediate based on an analysis of the burst-phase signal (3.5 kcal mol^{-1} and 1.3 kcal mol^{-1} M^{-1}).[10] The lower m-value compared to wild type (1.2 versus 2.1 kcal mol^{-1} M^{-1}) indicates that less solvent accessible surface area is buried upon folding, consistent with the lowest energy conformation of these mutants being a partially unfolded species. The presence or absence of mutations in Strand I does not appear to have an affect on the stability or amount of solvent accessible surface area buried upon folding for *E. coli* RNase H.

The results for *T. thermophilus* RNase H variants, however, revealed a difference between the mimic with and without mutations in strand I. While these mimics have very

similar m-values (also 1.2 kcal mol^{-1} M^{-1} compared to 2.0 kcal mol^{-1} M^{-1} for wild type *T. thermophilus* RNase H), the free energy of unfolding of TFLwt1 is 5.3 kcal mol^{-1} while that of TFLmut1 is 3.8 kcal mol^{-1}, a difference of 1.5 kcal mol^{-1}.

In sum, in spite of destabilizing the native state by the total free energy of folding, the proteins are not entirely unfolded. All four show CD spectra that are inconsistent with either the native state or the unfolded state, show denaturant melts that can be fit with a two-state assumption with m-values lower than that expected for the folded protein. Together, these data suggest all four variants adopt partially-folded proteins.

Investigating the structural properties of the intermediate mimics by NMR

In attempt to obtain residue-specific information about the structural properties of the intermediate mimics, ^{15}N-^{1}H HSQC spectra were obtained on a 900 MHz spectrometer. A high magnetic field and long experiment times were required due to the low concentrations of the partially folded proteins (70 – 120 μM). All spectra were obtained in 20 mM sodium acetate, pH 5.5, 50 mM KCl, and 10 % D_2O at 25°C.

The spectrum of EFLwt1 has some dispersed peaks in addition to many collapsed (unfolded) peaks (Figure 4a). In order to compare this full-length *E. coli* RNase H intermediate mimic to a fragment model of the intermediate analogous to the one Bai and co-workers created of *T. thermophilus* RNase H, an HSQC of a fragment of *E. coli* RNase H corresponding to the folding core plus Strand I (named Efrag+1, the construct for which was a gift from L. Rosen) was also obtained. The spectra are very similar. Most of the peaks in the spectrum of Efrag+1 correspond to peaks in the spectra of the full-length mimic (Figure 4b). There are several additional dispersed peaks in the spectrum of the fragment; however, many of these become visible in the spectra of EFLwt1 when the signal threshold is significantly lowered (Figure 4b expansion). The two full-length *E. coli* RNase H mimics, with and without mutations in Strand I, have virtually identical spectra with peaks that overlay almost exactly (Figure 4c). (In Figure 4c, the spectra of EFLwt1 is overlaid with a that of S12G/P17A/I25A/R27A *E. coli* RNase H, the original mimic designed prior to taking the role of Strand I into consideration. Though it contains different mutations than EFLmut1, it does have mutations at two residues, S12 in Strand I and P17 in the turn after Strand I, that are included in the fragment retaining wild type Strand I.)

The spectrum of TFLwt1, shown in Figure 5a, also shows dispersed peaks in addition to a large set of collapsed (unfolded region) peaks. The spectrum is very similar to the Bai lab's fragment of *T. thermophilus* RNase H that includes Strand I[30], with the exception of the expected additional unfolded peaks in the middle of the spectra for the full-length intermediate mimics (Figure 5b). The two *T. thermophilus* RNase H mimics have very similar spectra, with no distinct differences between them (Figure 5c).

4.4 DISCUSSION

Protein engineering to model kinetic intermediates

By selectively destabilizing the native state, we generated partially folded species with properties similar to the kinetic intermediates of *E. coli* and *T. thermophilus* RNases H. The CD spectra (Figure 2) of the four full-length mimics are very similar in shape to that of the acid state molten globule of *E. coli* RNase H, whose structure mirrors that of its kinetic intermediate and a partially unfolded form detected in native state hydrogen

exchange experiments.[10] Furthermore, the stabilities and m-values are very close to the values calculated for the kinetic intermediate.

Dynamic properties of the *E. coli* RNase H folding intermediate

We recently reported that a single mutation in the periphery of *E. coli* RNase H, I25A, resulted in the equilibrium population of the intermediate[29]. The absence of signal from the folded regions of the intermediate in a series of HSQC's led us to postulate that the intermediate is a broad ensemble of interconverting states, dynamic on the timescale of NMR detection and thereby remaining concealed in the noise. The more complete population of the intermediate in the mutants used in this study allowed us to explore this idea more rigorously.

If the folded region of the *E. coli* RNase H intermediate mimic is indeed dynamic on a timescale that leads to significant line broadening, we would expect to observe only peaks from the unfolded regions of the intermediate, as we did with I25A. The spectra of the *E. coli* RNase H intermediate mimic EFLwt1 shows high-intensity peaks in the middle of the spectrum that correspond to unfolded protein (Figure 4a). There are, however, additional dispersed peaks that cannot be accounted for by unfolded protein. It is difficult to ascertain if there are enough folded peaks to account for the entire folded region, but there are certainly dispersed peaks that we did not observe in any of the I25A spectra. It is possible that our hypothesis that a broad ensemble of conformers in millisecond conformational exchange is incorrect. If so, why were these peaks undetectable in the spectra of I25A? In the absence of denaturant, the native state should account for 80 % of the molecules, and peaks from the structured region of the intermediate could be obscured by the signal from this highly populated state. At 2 M urea, 60 % of the molecules should populate the intermediate, and we expect to observe peaks from this species. The presence of denaturant, however, may bring the conformational exchange rates into the intermediate time regime, leading to broadening and loss of signal. The question of the dynamic properties of the folded region of the intermediate under native conditions remains unanswered.

A comparison of the corresponding fragment intermediate mimic of *E. coli* RNase H, the folding core plus Strand I (Efrag+1), may provide some insight into the dynamic properties of the intermediate. The HSQC of Efrag+1 does not share many of the unfolded peaks visible in the spectrum of EFLwt1 (Figure 4b). (It may not share any of them, though it is difficult to interpret the middle of the spectrum, where unfolded peaks run together, with certainty.) Most of the peaks in the fragment HSQC correspond to the more dispersed peaks of EFLwt1; however, the fragment HSQC shows additional well-dispersed peaks. This is especially apparent in the region where peaks from the Trp side chains appear. There are six Trp's in Efrag+1, but there are at least 10 peaks in this region. The entire spectrum of the fragment has far more than 101 peaks (the number of residues in this construct).

There are several possible explanations for this. It is possible that there are two conformations. If this is true, they must be in slow exchange in order to both be detected, placing them on different sides of a rate-limiting barrier. It is tempting to suggest that Strand I is structured on the native side and unstructured on the unfolded side of the folding barrier. Another explanation is that the fragment is proteolyzed, and the extra peaks simply arise from an even smaller fragment; however no such fragment is evident by SDS-PAGE or mass spectrometry. Interestingly, many of the additional peaks in the

Efrag+1 HSQC become apparent in the spectrum of EFLwt1 when the signal threshold is significantly lowered. This is shown for the Trp region in the expansion of Figure 4b but is true for the extra peaks in other regions of the spectrum as well. Whatever species the extra peaks represent, it is also present in the full-length intermediate mimic.

Lastly, the oligomeric state of the intermediate mimics is not yet known, and there is precedence to suggest that they might not be monomeric. A fragment that consists of helices A, B, C, D, and Strand IV is known to be a dimer with well-defined peaks in its HSQC[33]. (The revised model of the "core" includes Strand V in addition to these secondary structure elements.) It was proposed that dimerization lends the packing specificity that results in a well-dispersed HSQC while the monomer might not be nearly so rigid. Perhaps both EFLwt1 and Efrag+1 are dimers, giving rise to detectable folded signal, while I25A is a monomer. This would explain the apparent incongruity between the mutants. We are currently carrying out experiments to determine the oligomeric state of the mimics.

Dynamic properties of the *T. thermophilus* RNase H folding intermediate

The analysis of the *T. thermophilus* RNase H intermediate mimic is more straightforward, as we have a spectrum of a known fragment structure from the Bai lab for comparison. The HSQC of TFLwt1 has a set of collapsed, unfolded peaks as well as many dispersed peaks (Figure 5a). These dispersed peaks are virtually entirely accounted for in the spectrum of the fragment of *T. thermophilus* RNase H that includes Strand I (Figure 5b). We know this fragment to be monomeric and well packed. The core of the full-length intermediate mimic is likely very similar in structure. It remains to be seen whether a rigid core in *T. thermophilus* RNase H is in contrast to a more molten, dynamic core of the *E. coli* RNase H intermediate, but the possibility brings up interesting questions. If the cores of the two proteins do behave very differently, what is the sequence origin of this difference? How does it fit into the overall biophysical mechanism of thermostabilization?

What is the role of Strand I?

The HSQC of a full-length intermediate mimic that contains mutations in Strand I is virtually identical to the spectrum of EFLwt1 (Figure 4c). Furthermore, the equilibrium denaturation curves of the two full-length mimics result in the same stability and m-values. It is unlikely that Strand I plays a role in stabilizing the *E. coli* RNase H kinetic intermediate.

The HSQC of TFLmut1 is also highly similar to TFLwt1 (Figure 5c); thus, all of *T. thermophilus* RNase H intermediate mimics (TFLmut1, TFLwt1, and the fragment that includes Strand I) have nearly the same set of dispersed peaks. This suggests that the cores of TFLmut1 and TFLwt1 resemble that of the fragment, which is known to be well-packed. TFLwt1, however, is more stable than TFLmut1 by 1.5 kcal/mol, indicating that Strand I must pack against the core in TFLwt1 and contribute to the folded structure. We do not yet know which intermediate mimic is an accurate representation of the folding intermediate. TFLmut1, with Strand I unstructured, may be analogous to the folding intermediate, and TFLwt1, with Strand I making stabilizing contacts, may represent an intermediate on the native side of the rate-limiting transition state. As suggested above, *E. coli* RNase H may have an analogous intermediate on the other side of the rate-limiting barrier, with Strand I folded against the core. In this case, however, it would be higher in energy than our canonical folding intermediate, as it appears by equilibrium

denaturation to have the same stability. Whatever the case, mutations in Strand I have different affects on the *E. coli* RNase H intermediate versus the *T. thermophilus* RNase H intermediate.

Many issues, including the oligomeric states and potential proteolysis of the intermediate mimics, remain to be worked out before strong conclusions can be drawn. Nevertheless, this preliminary data has provided interesting directions to pursue.

4.5 MATERIALS AND METHODS

Construction and Purification of RNase H Variants

The variants of *E. coli* and *T. thermophilus* RNases H used in this study were generated via the QuikChange Mutagenesis protocol. All of the single site mutants (except I25A, described in reference [29]) expressed solubly and were purified accordingly[10].

Due to problems with proteolysis during the expression and purification of the intermediate mimics, resulting in a contaminant corresponding to a fragment of the desired proteins and thus difficult to purify away, C-terminal His_6 tags were introduced to the four mimics (EFLwt1, EFLmut1, TFLwt1, and TFLmut1) via QuikChange. Constructs corresponding to the intermediate mimics of *E. coli* RNase H, EFLwt1 and EFLmut1, expressed insolubly and were purified as follows. Cell pellets were lysed via sonication in 25 mM Tris, pH 8, 20 mM NaCl, 1 mM EDTA, and the resulting pellet was resuspended in 25 mM Tris, 2 % Triton by briefly sonicating. The detergent wash was repeated, and protein was extracted from the resulting pellet via sonication in 50 % acetonitrile, 0.2 % TFA. Following extraction, the soluble portion was diluted by a factor of two, filtered, and lyophilized. The lyophilized protein was solubilized in 25 mM Tris, pH 8, 150 mM NaCl, 20 mM imidazole (Ni Buffer A), loaded onto a nickel column, and eluted with 25 mM Tris, pH8, 150 mM NaCl, and 500 mM imidazole (Ni Buffer B). The purity of the eluted protein was assessed by SDS-PAGE, and the protein was additionally loaded onto a HiTrap Capto S column if necessary, following dialysis into 25 mM sodium acetate pH 5.5, 100 mM NaCl, 1 mM EDTA (S Buffer A). Protein was eluted in 25 mM sodium acetate pH 5.5, 1 M NaCl, 1 mM EDTA. If a significant portion of the lyophilized protein could not be solubilized in Ni Buffer A, several milliliters of concentrated urea was added and the mixture was allowed to stir at room temperature. This mixture was then diluted with Ni Buffer A, dialyzed into Ni Buffer A, and run over the nickel and S columns as needed.

The *T. thermophilus* RNase H proteins were present in both the soluble and insoluble fractions; these proteins were purified from the insoluble portion as described above. The soluble fraction was loaded directly onto a nickel column in Ni Buffer A, and protein was eluted in Ni Buffer B. Protein was further subjected to an S column as above, if necessary.

^{15}N-labeled protein for NMR experiments was expressed by transferring log-phase cells to M9 medium containing ^{15}N ammonium chloride[34]. Mutants were purified as described above.

Equilibrium CD Measurements

Experiments were performed on an Aviv 410 circular dichroism spectropolarimeter in buffer containing 20 mM NaOAc, pH 5.5, and 50 mM KCl. Protein concentration was determined by extinction coefficient calculated according to

the number of Trp and Tyr residues in the protein[35]. CD spectra were collected in a 1 mm pathlength cuvette at a protein concentration of 500 ug/mL.

For urea denaturation experiments, samples in varying concentrations of urea containing 40 – 50 ug/mL protein were equilibrated overnight. CD signal was measured at 222 nm in a 1 cm pathlength cuvette at 25 ^{o}C. The signal over 60 seconds was averaged for each sample.

NMR Experiments

Two-dimensional ^{15}N-^{1}H HSQC's were recorded on a Bruker Avance II 900 MHz spectrometer equipped with a TCI cryoprobe at 25^{o}C. 96 scans were collected with 512 points in the direct dimension and 200 complex points in the indirect dimension. These data were processed using NMRPipe and viewed in CARA.

4.6 ACKNOWLEDGEMENTS
We thank Jeff Pelton and David Wemmer for NMR use and guidance. This work was supported by NIH grant GM50945 to S.M.

4.7 REFERENCES

1. Clark, A. C. (2008). Protein folding: are we there yet? *Arch Biochem Biophys* **469**, 1-3.
2. Baldwin, R. L. (2008). The search for folding intermediates and the mechanism of protein folding. *Annu Rev Biophys* **37**, 1-21.
3. Jahn, T. R. & Radford, S. E. (2005). The Yin and Yang of protein folding. *FEBS J* **272**, 5962-70.
4. Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H. & Clore, G. M. (1993). Kinetics of folding of the all-beta sheet protein interleukin-1 beta. *Science* **260**, 1110-3.
5. Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-61.
6. Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* **224**, 771-82.
7. Miller, E. J., Fischer, K. F. & Marqusee, S. (2002). Experimental evaluation of topological parameters determining protein-folding rates. *Proc Natl Acad Sci U S A* **99**, 10359-63.
8. Dalessio, P. M. & Ropson, I. J. (2000). Beta-sheet proteins with nearly identical structures have different folding intermediates. *Biochemistry* **39**, 860-71.
9. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192-7.
10. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* **4**, 298-304.
11. Teilum, K., Poulsen, F. M. & Akke, M. (2006). The inverted chevron plot measured by NMR relaxation reveals a native-like unfolding intermediate in acyl-CoA binding protein. *Proc Natl Acad Sci U S A* **103**, 6877-82.
12. Teilum, K., Maki, K., Kragelund, B. B., Poulsen, F. M. & Roder, H. (2002). Early kinetic intermediate in the folding of acyl-CoA binding protein detected by fluorescence labeling and ultrarapid mixing. *Proc Natl Acad Sci U S A* **99**, 9807-12.
13. Neudecker, P., Lundstrom, P. & Kay, L. E. (2009). Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys J* **96**, 2045-54.
14. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **286**, 1597-608.
15. Friel, C. T., Beddard, G. S. & Radford, S. E. (2004). Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J Mol Biol* **342**, 261-73.
16. Spudich, G. M., Miller, E. J. & Marqusee, S. (2004). Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol* **335**, 609-18.

17. White, G. W., Gianni, S., Grossmann, J. G., Jemth, P., Fersht, A. R. & Daggett, V. (2005). Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J Mol Biol* **350**, 757-75.
18. Neuweiler, H., Doose, S. & Sauer, M. (2005). A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate. *Proc Natl Acad Sci U S A* **102**, 16650-5.
19. Mogensen, J. E., Ipsen, H., Holm, J. & Otzen, D. E. (2004). Elimination of a misfolded folding intermediate by a single point mutation. *Biochemistry* **43**, 3357-67.
20. Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H. & Morikawa, K. (1993). Crystal structure of ribonuclease H from Thermus thermophilus HB8 refined at 2.8 A resolution. *J Mol Biol* **230**, 529-42.
21. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from E. coli. *Nature* **347**, 306-9.
22. Hollien, J. & Marqusee, S. (2002). Comparison of the folding processes of T. thermophilus and E. coli ribonucleases H. *J Mol Biol* **316**, 327-40.
23. Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* **6**, 825-31.
24. Connell, K. B., Miller, E. J. & Marqusee, S. (2009). The folding trajectory of RNase H is dominated by its topology and not local stability: a protein engineering study of variants that fold via two-state and three-state mechanisms. *J Mol Biol* **391**, 450-60.
25. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057-60.
26. Hollien, J. & Marqusee, S. (1999). A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* **38**, 3831-6.
27. Robic, S., Berger, J. M. & Marqusee, S. (2002). Contributions of folding cores to the thermostabilities of two ribonucleases H. *Protein Sci* **11**, 381-9.
28. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M. & Marqusee, S. (2003). Role of residual structure in the unfolded state of a thermophilic protein. *Proc Natl Acad Sci U S A* **100**, 11345-9.
29. Connell, K. B., Horner, G. A. & Marqusee, S. (2009). A single mutation at residue 25 populates the folding intermediate of E. coli RNase H and reveals a highly dynamic partially folded ensemble. *J Mol Biol* **391**, 461-70.
30. Zhou, Z., Feng, H., Ghirlando, R. & Bai, Y. (2008). The high-resolution NMR structure of the early folding intermediate of the Thermus thermophilus ribonuclease H. *J Mol Biol* **384**, 531-9.
31. Hollien, J. & Marqusee, S. (1999). Structural distribution of stability in a thermophilic enzyme. *Proc Natl Acad Sci U S A* **96**, 13674-8.
32. Pace, C. N. & Shaw, K. L. (2000). Linear extrapolation method of analyzing solvent denaturation curves. *Proteins* **Suppl 4**, 1-7.

33. Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Sci* **8**, 2251-7.
34. Marley, J., Lu, M. & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-5.
35. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948-54.

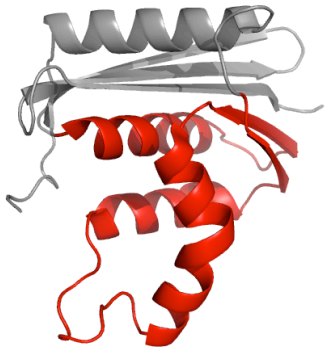
Table 4.1: A summary of the mutations used to construct the *E. coli* RNase H intermediate mimics. A checkmark under the EFLwt1 and EFLmut1 columns indicates the presence of the mutation in that construct. The number in the bottom row, the additive ΔΔG, is the total destabilization to the native state of *E. coli* RNase H, assuming the mutations contribute additively. The stability of wild type *E. coli* RNase H is 9.7 kcal/mol.

Mutation	**Location**	**ΔΔG (kcal/mol)**	**EFLwt1**	**EFLmut1**
F8A	Strand 1	2.7		✓
S12G	Strand 1	0.9		✓
I25A	Strand 2	5.1	✓	
R27A	Strand 2	2.2	✓	✓
S36G	Strand 3	2.6	✓	✓
E135G	Helix E	1.5		✓
Additive ΔΔG			9.9	9.9

Table 4.2: A summary of the mutations used to construct the *T. thermophilus* RNase H intermediate mimics. A checkmark under the TFLwt1 and TFLmut1 columns indicates the presence of the mutation in that construct. The number in the bottom row, the additive ΔΔG, is the total destabilization to the native state of *T. thermophilus* RNase H, assuming the mutations contribute additively. The stability of wild type *T. thermophilus* RNase H is 12.8 kcal/mol.

Mutation	**Location**	**ΔΔG (kcal/mol)**	**TFLwt1**	**TFLmut1**
F8A	Strand 1	3.5		✓
A12G	Strand 1	1.2		✓
A23G	Strand 2	2.7	✓	✓
L25A	Strand 2	3.6	✓	
L34A	Strand 3	2.2	✓	✓
S36G	Strand 3	1.5	✓	✓
R135G	Helix E	1.5	✓	✓
A137G	Helix E	2.9	✓	
Additive ΔΔG			14.4	12.6

(a)



(b)

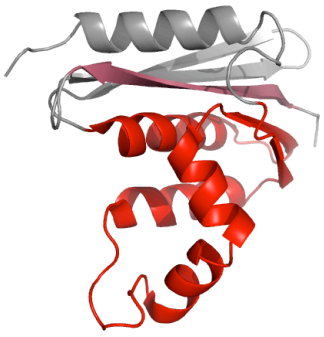


Figure 4.1: Ribbon structures of *E. coli* (a) and *T. thermophilus* (b) RNases H with the "core" regions colored red. Strand I in *T. thermophilus* RNase H, which was included in a fragment model of the kinetic intermediate, is colored pink.

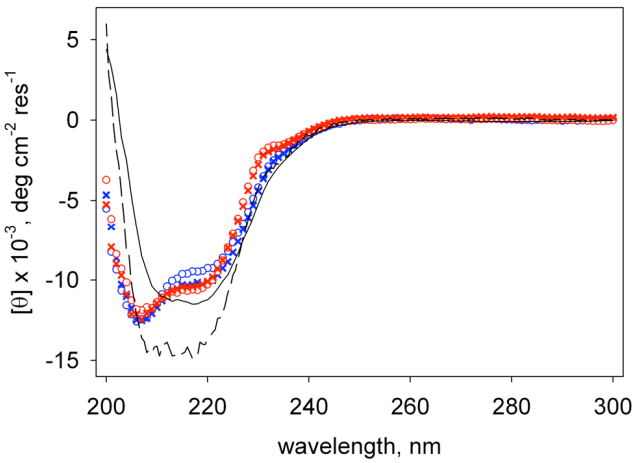
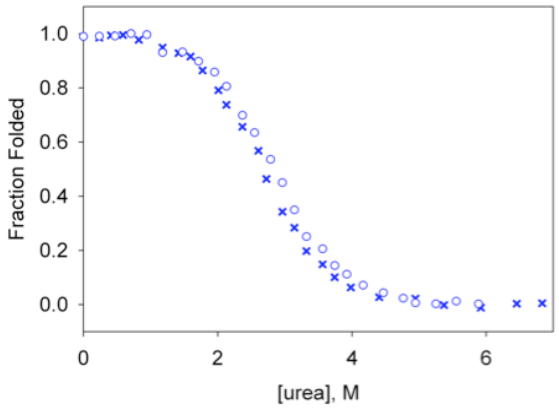


Figure 4.2: CD spectra of the full-length intermediate mimics. EFLwt1 (open blue circles), EFLmut1 (blue crosses), TFLwt1 (open red circles), and TFLmut1 (red crosses) are overlaid with wild type *E. coli* RNase H (solid line) and wild type *T. thermophilus* RNase H (dashed line).

(a)



(b)

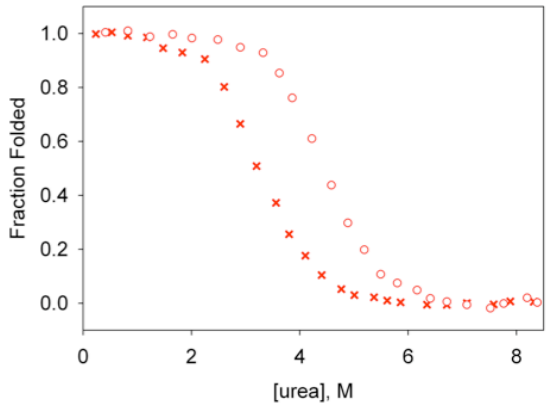


Figure 4.3: Equilibrium denaturant melts of the full-length *E. coli* (a) and *T. thermophilus* (b) intermediate mimics plotted as fraction folded. The mimics without mutations in Strand I, EFLwt1 in (a) and TFLwt1 in (b), are shown as open circles, and the mimics with mutations in Strand I, EFLmut1 in (a) and TFLmut1 in (b), are shown in crosses.

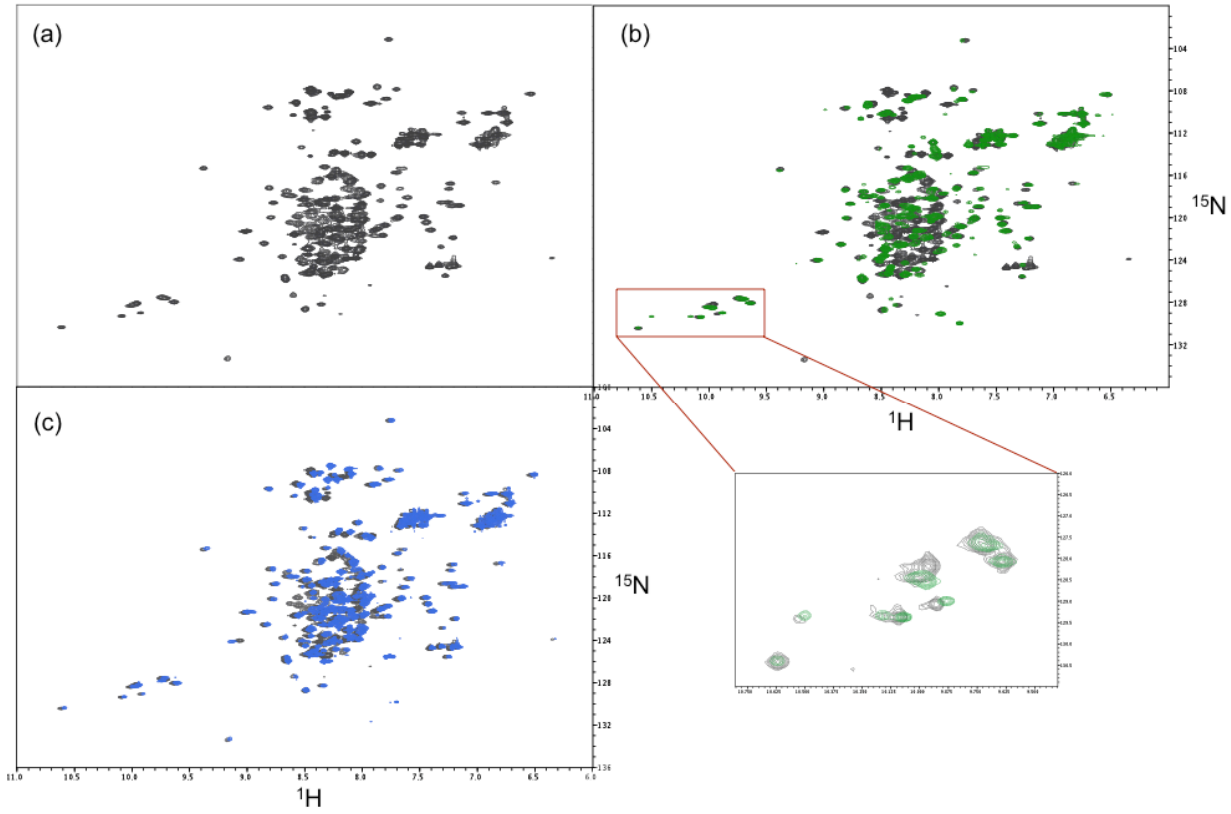


Figure 4.4: ^{15}N-^{1}H HSQC spectra of the *E. coli* RNase H intermediate mimics. (a) EFLwt1 shown by itself. (b) EFLwt1 in grey is overlaid with Efrag+1 in green. The region in the red box is expanded to show the Trp peaks at a reduced signal threshold. (c) EFLwt1 in grey overlaid with S12G/P17A/I25A/R27A *E. coli* RNase H in blue.

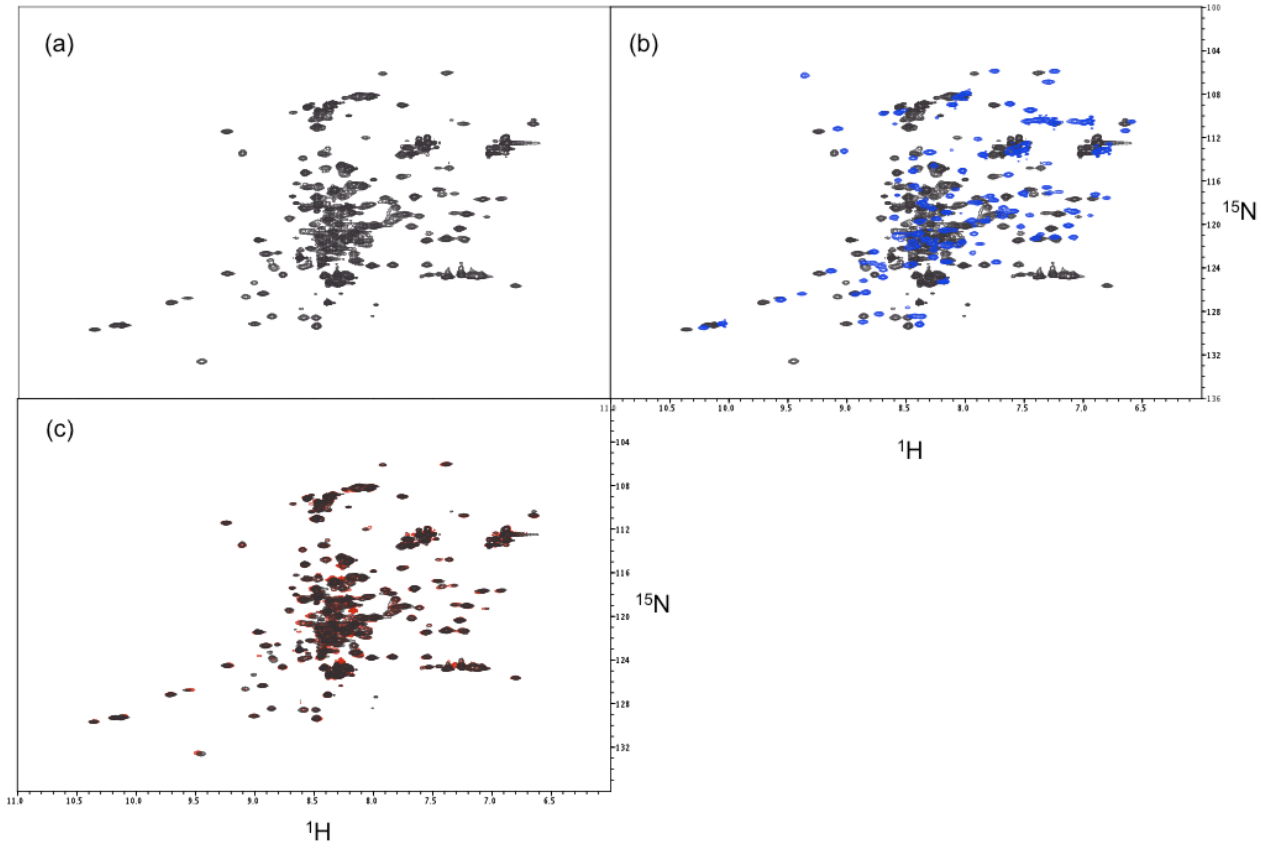


Figure 4.5: ^{15}N-^{1}H HSQC spectra of the *T. thermophilus* intermediate mimics. (a) TFLwt1 is shown by itself. (b) TFLwt1 in grey is overlaid with the HSQC of the fragment of RNase H including Strand I reported in reference 30 in blue. (c) TFLwt1 in grey overlaid with TFLmut1 in red.

Chapter V

Cooperativity under force: Single-molecule mechanical folding studies of an ankyrin repeat protein

5.1 ABSTRACT

Repeat proteins are ideal candidates for single-molecule mechanical unfolding studies given their modular architecture and their potential role as elastic elements in the cell. Several ankyrin repeat proteins have been studied in the atomic force microscope, where they have been observed to unfold in a largely stepwise manner, one repeat at a time. This is in direct contrast to the cooperative transitions observed in bulk experiments. To explore how a modular system builds cooperativity into its energetics, the limits of this cooperativity, and the origin of the different responses to force and chemical denaturant, we studied the behavior of the Notch ankyrin domain in the optical tweezers. Our results indicate that the forced-unfolding of the Notch ankyrin domain, a seven-ankyrin repeat protein, occurs in one or two steps when manipulated in the optical tweezers. Though the unfolding pathway is heterogeneous compared to that observed in bulk studies, there is a limit to the degree to which repeats can be uncoupled from one another. We compare our results to the unfolding behavior of the Notch ankyrin domain in the AFM obtained by our collaborators for this project (Doug Barrick and Naomi Courtemanche at Johns Hopkins University). This offers some insight into the apparent difference in solution and AFM unfolding of ankyrin repeat proteins. This work is ongoing, and here I present our initial results.

5.2 INTRODUCTION

Repeat proteins are a class of proteins containing a variable number of identical secondary structure motifs that stack in a roughly linear fashion to produce a modular architecture. In spite of the lack of long-range sequence contacts, many repeat proteins exhibit the cooperativity characteristic of globular domains, providing an interesting framework for investigating the mechanism by which distinct modules are energetically coupled. The simple tandem array of secondary structure units lends itself to studies in which the protein can be dissected and the equilibrium energy landscape mapped[1; 2]. It also makes them ideal candidates for single-molecule mechanical unfolding studies. Force can be applied along the repeat axis, which may result in a loss of coupling between repeats, thus providing a direct test of the limits of cooperativity. These experiments also have the advantage of tracking one molecule at a time[3], uncovering any heterogeneity masked by the averaging inherent to ensemble studies. Such heterogeneity would be expected for such structurally degenerate proteins.

Ankyrin repeats, a well-studied class of repeat proteins[4; 5; 6; 7], are comprised of 33 residues long repeats consisting of two alpha helices connected by a short loop. Ankyrin repeat proteins have been implicated in a diverse set of biological functions, such as cell-cell signaling, cytoskeletal integrity, and transcriptional regulation, with a unifying theme that appears to be the mediation of protein-protein interactions[8; 9]. Functions in which the mechanical response of the protein may play a critical role have also been proposed[10; 11]. Specifically, copies of ankyrin repeats have been found to be a component of transduction channels in mechanosensitive hair cells involved in hearing in vertebrates[12; 13]. In light of their possible involvement, there has been considerable interest in characterizing the behavior of ankyrin repeat proteins under force.

The forced-unfolding of three ankyrin repeat proteins has been studied in the atomic force microscope (AFM). N6C, an array of six consensus ankyrin repeats flanked by N- and C-terminal soluble caps, and ankyrin-B, a naturally-occurring 24-repeat

protein, exhibited stepwise unfolding behavior in the AFM[14; 15]. The contour length changes associated with the transitions correspond to the unfolding of one repeat at a time, indicating a complete decoupling of the individual repeats. Gankyrin, a natural ankyrin repeat protein containing seven repeats, was observed to maintain a limited amount of cooperativity between repeats[16]; although many transitions represented the unfolding of half a repeat or one repeat, unfolding events were detected that were larger than two repeats in size. It has been suggested that this stepwise response has biological significance: the repeats can act as force buffers whose ability to unfold one repeat at a time allow them to resist force over a range of extensions.

Such a loss of coupling between repeats is somewhat surprising, however, considering the cooperativity observed in bulk studies. Two-state equilibrium unfolding transitions have been observed for ankyrin repeat proteins containing as many as six repeats[5; 6; 7]. The high rate at which the force is applied in the AFM may be at least partly responsible for multi-state behavior. These high loading rates result in conditions far from equilibrium[17]. Domains of the cytoskeletal protein spectrin were also unfolded in a multi-state fashion in the AFM, in contrast to their cooperative unfolding transition observed in bulk kinetic experiments[18]. The optical tweezers, another single-molecule mechanical technique, are capable of applying force more slowly, allowing the system to stay closer to equilibrium[19; 20]. The lower range of forces accessible in the optical tweezers may also have more biological relevance, although the range of forces relevant for mechanical processes in the cell remain largely uncharacterized[21; 22; 23]. The 3.5 MDa protein titin has been characterized in both the AFM[24] and optical tweezers[25]; however the unfolding of individual globular domains was obscured by the size and heterogeneous composition of the molecule. A thorough comparison of the same protein domain in the high (AFM) and low (optical tweezers) force regimes is still lacking.

To further explore the limits of cooperativity of a modular system and in attempt to reconcile the differences in unfolding behaviors of ankyrin repeat proteins in response to force and chemical denaturant, we examined the folding of the Notch ankyrin domain in the optical tweezers. The Notch ankyrin domain is a roughly linear array of seven ankyrin repeats (Figure 1)[26]. Thermodynamic studies indicate that the Notch ankyrin domain exhibits a two-state equilibrium unfolding transition[5]. Bulk studies have explored the limits of this cooperativity, finding that the protein requires four repeats to fold and that the intrinsically unstable repeats are stabilized by the interfaces between them[27; 28]. The kinetic trajectory has also been characterized; the Notch ankyrin domain folds through a discrete pathway with an intermediate characterized by structure in the central repeats[29]. Here, we find that the protein can unfold in a variety of pathways in one or two steps; however, stepwise unfolding of single repeats is not observed. In addition, we find that the last step to unfolding is at least three repeats in size, indicating that the Notch ankyrin domain does maintain a level of cooperativity under force. Using a similar construct in the AFM, the Barrick lab finds that the unfolding of the Notch ankyrin domain is even more heterogeneous, and the size limit on the number of repeats capable of resisting force is smaller. A comparison of these techniques helps to explain the loss of observed cooperativity from bulk to single-molecule measurements.

5.3 RESULTS

Generation of the protein-DNA chimera

In order to study the Notch ankyrin domain in the optical tweezers, we first had to covalently attach the protein to DNA "handles" through cysteine chemistry[19; 30]. The handles provide the space necessary to manipulate the molecule, which is tethered between one bead held on the end of a micropipette and one held in the optical trap. Sulfo-SMCC, a heterobifunctional cross-linker, was used to link stretches of double-stranded DNA 558 base pairs in length to the protein molecule (see Materials and Methods for details). A schematic of the DNA handle attachment reaction is shown in Figure 2.

The Notch ankyrin domain unfolds in one or two steps under force

The DNA-protein-DNA complex was subjected to consecutive stretching and relaxing cycles in the optical tweezers at loading rates between 5 and 20 pN/s. Representative force-extension curves of the Notch ankyrin domain are shown in Figure 3. A single unfolding event was observed in 199 of 413 pulling cycles (48 %). Two unfolding events were detected in the other 214 curves. Both types of behavior were observed for every protein molecule that we were able to capture.

The change in contour length in a given transition provides important information about how much of the protein chain is liberated (i.e. unfolded) in that transition. The expected change in contour length for the unfolding of a single ankyrin repeat is calculated as follows:

(33 residues/repeat * 0.36 nm/unfolded residue) - 0.8 nm/folded repeat = 11.1 nm

The first repeat of the Notch ankyrin domain does not appear to be well-folded and is not expected to contribute to a change in contour length upon unfolding. This repeat adopts three different conformations in the three molecules in the unit cell of the crystal structure,[31] and introducing a mutation in repeat 1 does not alter stability[27]. Therefore, when the entire Notch ankyrin domain unfolds, we expect a contour length change arising from the unfolding of the six ordered repeats, 66.6 nm.

One unfolding transition

When one unfolding event is detected in a force-extension cycle (48% of the time), the contour length change most often accounts for the unfolding of the entire protein, as shown by the histogram in Figure 4a. In spite of its modularity, when pulled at loading rates of 5 – 20 pN/s, the protein is capable of unfolding in an all-or-none fashion under force. There also appear to be some single transitions of ~44 and 56 nm, corresponding to the unfolding of four and five repeats. (Given the noise in these measurements and the manual detection of unfolding transitions, we believe differences of up to a few nm from the expected values is reasonable.) It is highly unlikely that one or two repeats remain folded at the upper force; therefore, these transitions likely indicate that the protein does not always completely refold as it is relaxed.

Two unfolding transitions

For the 52% of the pulls that resulted in two discernable transitions, the histograms of the contour length changes associated with the first and second unfolding transitions are shown in Figures 4b and 4c. The first unfolding event has two peaks that match the expected change for two and three repeats (Figure 4b). The amount of protein unfolded in the second event is harder to interpret (Figure 4c), as the distribution peaks around 30 nm and is marked by the presence of events larger in size. To obtain a better understanding of the unfolding trajectory when two events are detected, the first and second events in a given curve showing two transitions were summed (Figure 5a) and

subtracted (Figure 5b). The total contour length change is centered at approximately 64 nm when both transitions are summed, confirming that the two events account for the unfolding of the entire protein. In Figure 5b, the contour length change of the first event was subtracted from that of the second. This distribution has its largest peak around zero nm, indicating that most often the Notch ankyrin domain unfolds in two steps of three repeats. Strikingly, there are many pulling cycles for which the change in contour length of the second event is larger than that of the first, but not vice versa, reflected by the abundance of events with a positive difference and the relative absence of events with a negative difference. Therefore, when the two unfolding events are not equal in size, a greater portion of the protein remains structured after the first unfolding event. There are small peaks (~13 and 23 nm) that may correspond to a difference of one and two repeats between the two transitions.

Evidence for a refolding intermediate

Refolding events are difficult to detect in the optical tweezers due to the increased noise and small changes in extension expected at low force. Insight into the refolding pathway, however, can be gained indirectly by varying the time spent waiting at 1 pN between stretching and relaxing cycles and examining the subsequent unfolding event(s). If the protein folds through an intermediate, as the incubation time at 1 pN decreases (the protein has less time to completely refold), we expect to observe an increase in unfolding transitions that represent the unfolding of an intermediate state rather than the fully folded state.

Figure 6 displays histograms representing the total contour length change associated with unfolding events after waiting 5, 3, and 1 seconds at 1 pN. For the shorter pause duration pulls, we observe an increase in the number of pulls in which the entire protein is not accounted for in the total contour length change associated with the unfolding event(s). Specifically, the histograms reflect an increase in the percentage of unfolding events corresponding to four repeats. The protein has less time to fully refold, and this represents the unfolding of a refolding intermediate of four structured repeats. As the waiting time is decreased, there is not enough time for the protein to even partially refold and fewer curves show any transitions at all, which contributes to the deterioration in the quality of the histograms. Still, it is clear that the proportion of observed events that correlate to the unfolding of four repeats increases as refolding time decreases.

5.4 DISCUSSION

Unfolding pathways of the Notch ankyrin domain under force

The Notch ankyrin domain unfolds in one or two transitions when manipulated by force using the optical tweezers. The single unfolding event indicates that in spite of its modularity, the entire protein can unfold in one cooperative transition. When the protein unfolds in two steps, it does so most often in two equal steps of three repeats. There is at least one alternative route sampled, however, because the two transitions are not always equal in size. Therefore, there are at least three different unfolding routes observed under force: one in which the entire protein unfolds in one step, one in which three repeats unfold in each of two steps, and one which accesses an unfolding intermediate with more than three repeats remaining folded. The presence of a peak arising from the unfolding of two repeats in the first of two unfolding transitions and a peak corresponding to a

difference of two repeats between the two events strongly suggests the presence of a four-repeat unfolding intermediate.

These data cannot tell us exactly which repeats unfold in each transition or if the same repeats always unfold first, and it is possible that there are many more than three routes to the unfolded state. Repeats may unfold from one or both termini or from the central repeats. Unfolding from the termini seems more plausible for several reasons. Unfolding from terminal repeats disrupts one interface between repeats while unfolding from the middle of the domain would disrupt two repeat interfaces. Moreover, this route would have an unfolding intermediate with two units of structure separated by unfolded chain that is stable to force. Lastly, it is known from work in bulk that the central repeats are the most stable[28]. Repeat 6 is known to be intrinsically unstable and in fact does not remain folded in the absence of repeat 7. It therefore seems likely if repeat 7 were to unfold, repeat 6 would do so without the application of additional force.

Given the broad distribution of contour length changes, it is also possible that the Notch ankyrin domain unfolds in a non-integral number of repeats. A single helix of a repeat may unfold while the other maintains contacts with the remainder of the folded stack. In fact, the only pathway that could result in a difference of one repeat but still unfold all six is one in which 2.5 repeats unfold first, followed by 3.5 in the second step.

As for the four-repeat refolding intermediate, its structure may resemble that observed in bulk, which is characterized by structure in repeats 3-5 and to a lesser extent, repeat 2[29].

Cooperativity under force

There is no evidence for parallel unfolding or refolding pathways for the Notch ankyrin domain in bulk[29] This difference in heterogeneity likely arises from the application of force over a defined axis versus the use of chemical denaturant, which acts uniformly over the molecule. In contrast to denaturation by chemicals, a partial unfolding event under force can alleviate strain on the molecule, thereby reducing the destabilization to the remaining folded units. This effect can be achieved through a number of different pathways in a modular protein and thus contributes to the observed increase in heterogeneity.

Compared to bulk studies, the individual repeats can respond to force in a greater number of ways, indicating some degree of decoupling of the units (i.e. decreased cooperativity). The limit on the extent of this decoupling, however, is also apparent. For a large number of the force-extension curves collected, the Notch ankyrin domain acts as a cooperative unit and unfolds in a single transition. When it does not, we do not observe the stepwise unfolding of single repeat units or a final unfolding transition of less than three repeats. When the protein unfolds in more than one transition, the size of the last event is a measure of the smallest number of units that is kinetically stable to force. For the Notch ankyrin domain, this size limit appears to be three repeats. One and two-repeat units are not independently stable to the force required to initiate unfolding.

Comparison of force regimes: optical tweezers versus AFM

The cooperativity of repeat proteins can also be explored by comparing optical tweezer data to atomic force microscopy (AFM) data. The two single-molecule mechanical techniques access different force regimes. The lower spring constant of the optical trap results in a much lower loading rate, the speed at which force is applied to the molecule (low force regime). The AFM applies force much more quickly (high force

regime) and thus brings the system farther out of equilibrium. Indeed, the response of the Notch ankyrin domain to force in the AFM uncovers additional types of unfolding behavior. While one and two unfolding transitions are observed in the optical tweezers, three sequential events are detected in the AFM, and this appears to correspond to the unfolding of two repeats at a time (Courtemanche and Barrick, unpublished results). Even at these high loading rates, however, the protein unfolded in a single transition in ~35 % of the curves. Though the size of the last transition is heterogeneous and difficult to interpret, it is clear that the threshold of stability is greater than one repeat but smaller than the three repeats observed in the tweezers. As the system is moved farther out of equilibrium, it behaves less cooperatively. The intermediates observed in these studies may represent meta-stable species that would not relax to an energy minimum during the slower pulling. The forced-unfolding studies uncover complexities in a rough energy landscape that are inaccessible by bulk chemical denaturation. At this point, the biological implications this changing response to force might have for ankyrin repeat proteins are unclear. Further characterization of the role ankyrin repeat proteins might have in mechanical processes in the cell may shed light on what force regime is most relevant for these functions.

Comparison to mechanical unfolding studies of other ankyrin repeat proteins

The Notch ankyrin domain demonstrates markedly different behavior from other ankyrin repeat proteins in response to force. Both N6C and ankyrin-B exhibit multiple, stepwise, low-force transitions around 40 - 50 pN in the AFM that arise from the unfolding of one repeat. Slightly different behavior was observed for gankyrin, a natural seven-ankyrin repeat protein. While most events correspond to the unfolding of half or one full repeat, the unfolding of two or more was also frequently observed. The lower force regime of the optical tweezers is one possible reason for the difference, although the Notch ankyrin domain in the AFM still shows fewer unfolding transitions, lacking the stepwise unfolding of single repeats at a time. Differences inherent to the various sequences of the proteins must also play a role. For all three proteins mentioned above, there is evidence for parallel pathways in bulk while the Notch ankyrin domain folds through a discrete pathway. The relevance these different responses to force may have for the function of ankyrin repeat proteins is unclear.

5.5 MATERIALS AND METHODS

Protein Expression and Purification

A plasmid containing a His_6-tagged construct of the Notch ankyrin domain with cysteines engineered at positions 13 and 268 was provided by Doug Barrick and Naomi Courtemanche. The plasmid was transformed into BL21 (DE3) pLysS cells, and cells were grown at 37 ^{o}C. At an O.D. ~0.6, cells were induced with 1 mM IPTG and grown for 3 – 4 more hours. Cells were harvested, lysed via sonication, and the soluble portion was purified via Ni^{2+}, anion exchange, and gel filtration chromatography as described[26]. All buffers contained 0.1 mM TCEP. Fractions containing pure protein were dialyzed into 25 mM Tris, pH 8.0, 150 mM NaCl, 0.1 mM TCEP and flash-frozen.

Optical Tweezer Sample Preparation

An analogous approach to the one described in Cecconi et al.[30] was used to generate the tweezer sample. In the original method, 2',2'-dithiodipyridine was used to functionalize the protein for subsequent attachment to the DNA handles. For unknown

reasons, the Notch ankyrin domain was refractory to this approach, and an alternative linker, the heterobifunctional sulfo-SMCC (Pierce), was utilized. Sulfo-SMCC is functionalized with an NHS ester at one end, which reacts with primary amines, and a maleimide at the other, which reacts with cysteines. The handle attachment approach is schematized in Figure 2 and outlined below.

Preparation of DNA Handles

DNA handles were designed to react with the NHS ester portion of the linker. Primers with the following 5'-end functionalities were ordered from IDT: a free amine, a biotin molecule, and a digoxigenin molecule. The 558-base pair pieces of DNA were generated via PCR using the amine-linked primer and either the biotin or digoxigenin-linked primer and the pGEMEX 1 plasmid as a template. The PCR product was purified using a Qiagen maxi-prep kit. In the final step, DNA was eluted from the column in sodium phosphate buffer, pH 7.4, concentrated, and frozen.

DNA Handle Attachment to Protein

1 mg of Sulfo-SMCC was dissolved in 50 μL DMSO. Approximately 0.1 nmol of each handle (Bio-DNA-NH_2 and Dig-DNA-NH_2) were incubated separately with 12.5 μL (500 nmol) of the sulfo-SMCC at 37 ^oC for 1 hour. Excess sulfo-SMCC was then removed via three consecutive buffer exchange columns containing Sephadex G-25 resin (Pierce) equilibrated in 100 mM sodium phosphate, pH 7.4. The resulting DNA was again concentrated. Meanwhile, an aliquot of the purified Notch ankyrin domain was buffer exchanged out of reducing agent via two consecutive Sephadex G-25 columns equilibrated in 100 mM phosphate, pH 7.4. The resulting protein was reacted with the DNA handles in a ratio of 4:1 total DNA:protein (2:1 for each handle:protein) at room temperature overnight. The reaction was assessed via 4-20% gradient SDS-PAGE.

Optical Tweezer Force Spectroscopy

Polystyrene beads coated in streptavidin were flowed into the tweezer chamber, and a single bead was caught by suction onto the end of a micropipette. A dilution of the reaction described above was incubated with polystyrene beads coated in anti-digoxigenin moieties for 15 minutes at room temperature. Individual antidigoxigenin-coated beads were caught in the laser trap. The Notch ankyrin domain was tethered between two polystyrene beads by bringing the two beads into close proximity. The protein undergoes stretching and relaxing cycles by moving the micropipette relative to the trap, and force is measured by the change in momentum of the light[20]. To ensure that the data reflects the pulling of a single fiber, each time a molecule is caught, a DNA overstretching transition is obtained and analyzed for the proper extension change. Data were collected at pulling speeds from 50 – 300 nm/s.

Analysis of Force-Extension Curves

Each force-extension curve was analyzed in LabView, and transitions were picked manually. The raw extension changes associated with each transition were converted into changes in contour length using the worm-like chain model:

$$F = \frac{k_B T}{p} \times \left[\frac{1}{4}\left(1 - \frac{x}{L_c}\right)^2 - \frac{1}{4} + \frac{x}{L_c} \right]$$

where F is the observed force, x is the chain extension, k_B is the Boltzmann constant, T is temperature, p is persistence length, and L_c is contour length.

REFERENCES

1. Kloss, E., Courtemanche, N. & Barrick, D. (2008). Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* **469**, 83-99.
2. Aksel, T. & Barrick, D. (2009). Analysis of repeat-protein folding using nearest-neighbor statistical mechanical models. *Methods Enzymol* **455**, 95-125.
3. Borgia, A., Williams, P. M. & Clarke, J. (2008). Single-molecule studies of protein folding. *Annu Rev Biochem* **77**, 101-25.
4. Tang, K. S., Guralnick, B. J., Wang, W. K., Fersht, A. R. & Itzhaki, L. S. (1999). Stability and folding of the tumour suppressor protein p16. *J Mol Biol* **285**, 1869-86.
5. Zweifel, M. E. & Barrick, D. (2001). Studies of the ankyrin repeats of the Drosophila melanogaster Notch receptor. 2. Solution stability and cooperativity of unfolding. *Biochemistry* **40**, 14357-67.
6. Mosavi, L. K., Williams, S. & Peng Zy, Z. Y. (2002). Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein. *J Mol Biol* **320**, 165-70.
7. Lowe, A. R. & Itzhaki, L. S. (2007). Biophysical characterisation of the small ankyrin repeat protein myotrophin. *J Mol Biol* **365**, 1245-55.
8. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* **13**, 1435-48.
9. Li, J., Mahajan, A. & Tsai, M. D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* **45**, 15168-78.
10. Howard, J. & Bechstedt, S. (2004). Hypothesis: a helix of ankyrin repeats of the NOMPC-TRP ion channel is the gating spring of mechanoreceptors. *Curr Biol* **14**, R224-6.
11. Gao, M., Sotomayor, M., Villa, E., Lee, E. H. & Schulten, K. (2006). Molecular mechanisms of cellular mechanics. *Phys Chem Chem Phys* **8**, 3692-706.
12. Walker, R. G., Willingham, A. T. & Zuker, C. S. (2000). A Drosophila mechanosensory transduction channel. *Science* **287**, 2229-34.
13. Sidi, S., Friedrich, R. W. & Nicolson, T. (2003). NompC TRP channel required for vertebrate sensory hair cell mechanotransduction. *Science* **301**, 96-9.
14. Li, L., Wetzel, S., Pluckthun, A. & Fernandez, J. M. (2006). Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. *Biophys J* **90**, L30-2.
15. Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V. & Marszalek, P. E. (2006). Nanospring behaviour of ankyrin repeats. *Nature* **440**, 246-9.
16. Serquera, D., Lee, W., Settanni, G., Marszalek, P. E., Paci, E. & Itzhaki, L. S. Mechanical unfolding of an ankyrin repeat protein. *Biophys J* **98**, 1294-301.
17. Ng, S. P., Randles, L. G. & Clarke, J. (2007). Single molecule studies of protein folding using atomic force microscopy. *Methods Mol Biol* **350**, 139-67.
18. Randles, L. G., Rounsevell, R. W. & Clarke, J. (2007). Spectrin domains lose cooperativity in forced unfolding. *Biophys J* **92**, 571-7.

19. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057-60.
20. Moffitt, J. R., Chemla, Y. R., Smith, S. B. & Bustamante, C. (2008). Recent advances in optical tweezers. *Annu Rev Biochem* **77**, 205-28.
21. Bustamante, C., Chemla, Y. R., Forde, N. R. & Izhaky, D. (2004). Mechanical processes in biochemistry. *Annu Rev Biochem* **73**, 705-48.
22. Hormeno, S. & Arias-Gonzalez, J. R. (2006). Exploring mechanochemical processes in the cell with optical tweezers. *Biol Cell* **98**, 679-95.
23. Oberhauser, A. F. & Carrion-Vazquez, M. (2008). Mechanical biochemistry of proteins one molecule at a time. *J Biol Chem* **283**, 6617-21.
24. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. (1997). Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-12.
25. Kellermayer, M. S., Smith, S. B., Granzier, H. L. & Bustamante, C. (1997). Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112-6.
26. Zweifel, M. E. & Barrick, D. (2001). Studies of the ankyrin repeats of the Drosophila melanogaster Notch receptor. 1. Solution conformational and hydrodynamic properties. *Biochemistry* **40**, 14344-56.
27. Bradley, C. M. & Barrick, D. (2002). Limits of cooperativity in a structurally modular protein: response of the Notch ankyrin domain to analogous alanine substitutions in each repeat. *J Mol Biol* **324**, 373-86.
28. Mello, C. C. & Barrick, D. (2004). An experimentally determined protein folding energy landscape. *Proc Natl Acad Sci U S A* **101**, 14102-7.
29. Bradley, C. M. & Barrick, D. (2006). The notch ankyrin domain folds via a discrete, centralized pathway. *Structure* **14**, 1303-12.
30. Cecconi, C., Shank, E. A., Dahlquist, F. W., Marqusee, S. & Bustamante, C. (2008). Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J* **37**, 729-38.
31. Zweifel, M. E., Leahy, D. J., Hughson, F. M. & Barrick, D. (2003). Structure and stability of the ankyrin domain of the Drosophila Notch receptor. *Protein Sci* **12**, 2622-32.

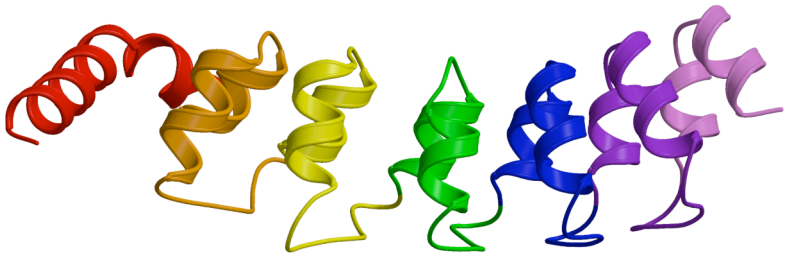


Figure 5.1: The crystal structure of the Notch ankyrin domain from *D. melanogaster* (ref) with the repeats 1 (red) to 7 (violet) separated by color. Repeat 1 is not expected to contribute to the total contour length chain upon unfolding (see text).

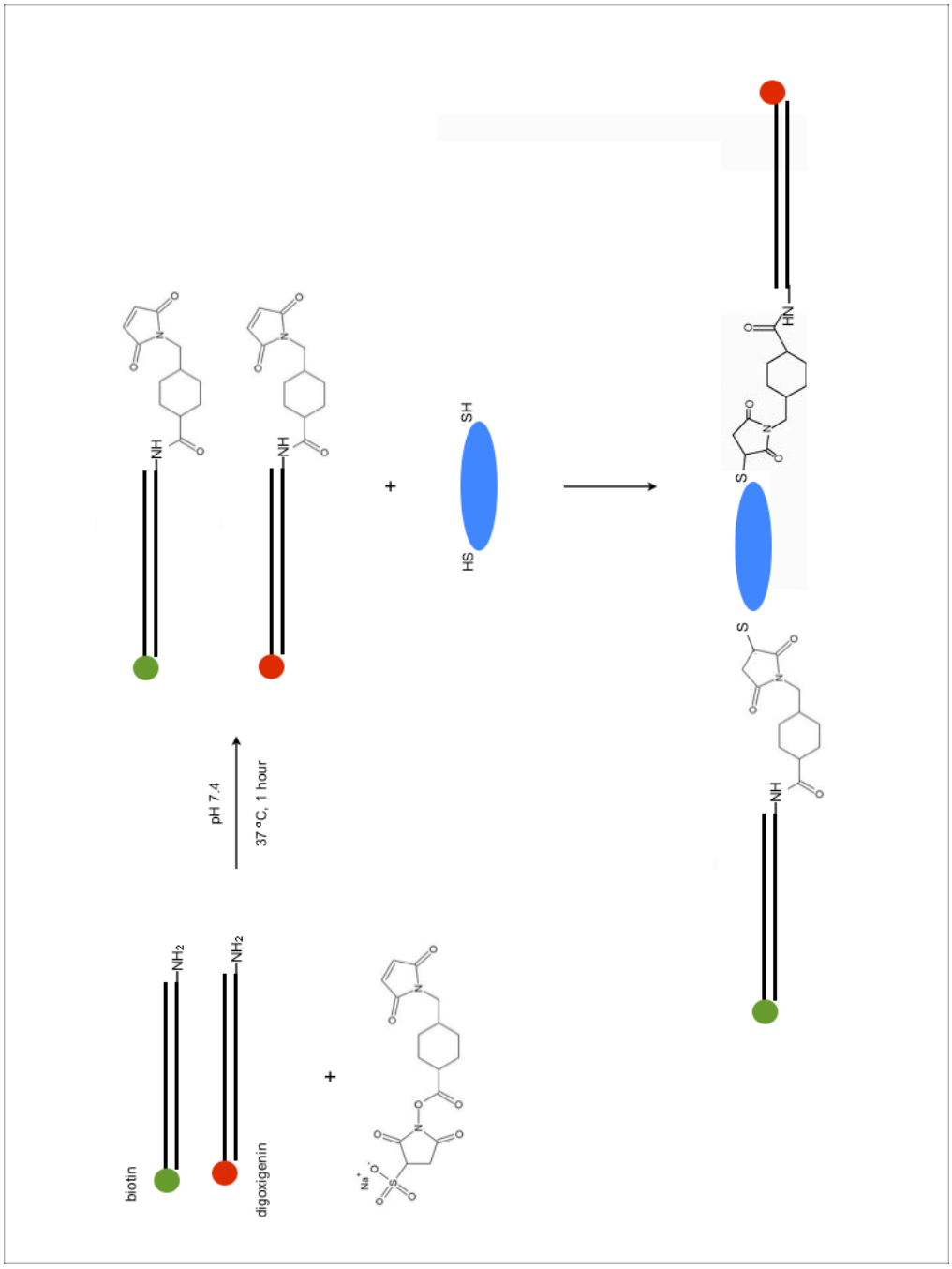


Figure 5.2: Schematic of DNA handle attachment to the protein. Sulfo-SMCC was reacted with DNA handles functionalized with either biotin or digoxigenin and a free

amine. After excess sulfo-SMCC was removed, these handles were reacted with protein with unique cysteines (blue oval in schematic) at the desired attachment points.

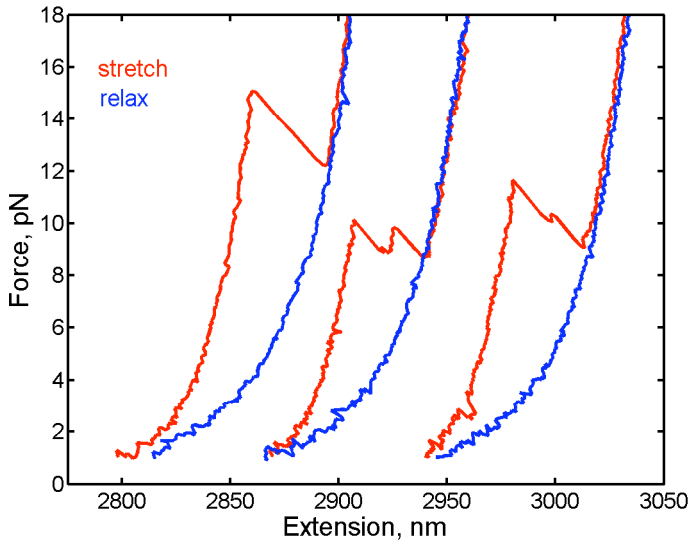
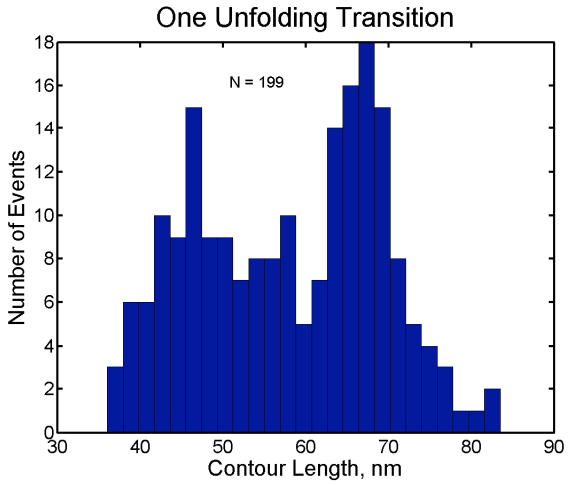
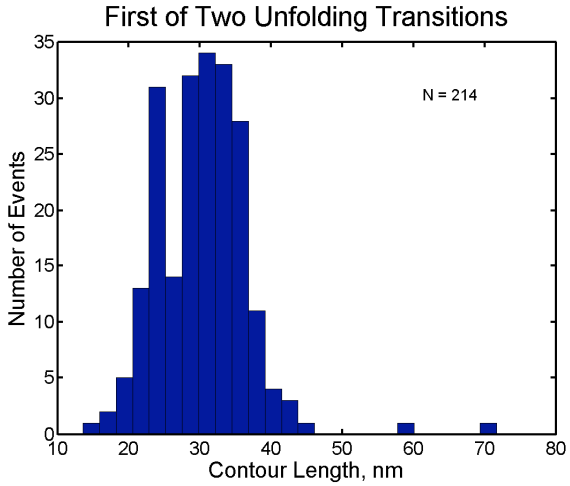


Figure 5.3: Representative force-extension curves of the Notch ankyrin domain. The protein unfolded in one (left curve) or two (center and right curves) transitions in the optical tweezers.

(a)


One Unfolding Transition
N = 199
Number of Events
Contour Length, nm

(b)


First of Two Unfolding Transitions
N = 214
Number of Events
Contour Length, nm

(c)

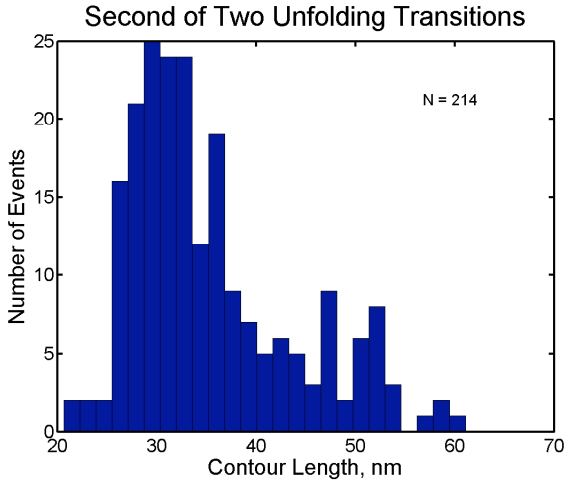
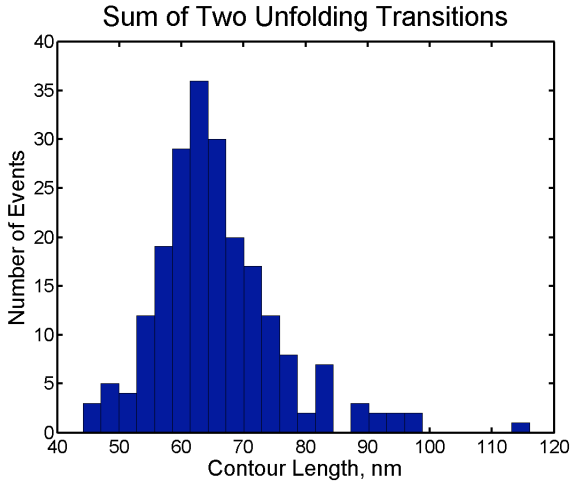

Second of Two Unfolding Transitions
N = 214
Number of Events
Contour Length, nm

Figure 5.4: Distribution of the contour length changes of the Notch ankyrin domain associated with (a) a single unfolding transition and (b) the first and (c) second of two unfolding transitions. N is the total number of events represented.

(a)



(b)

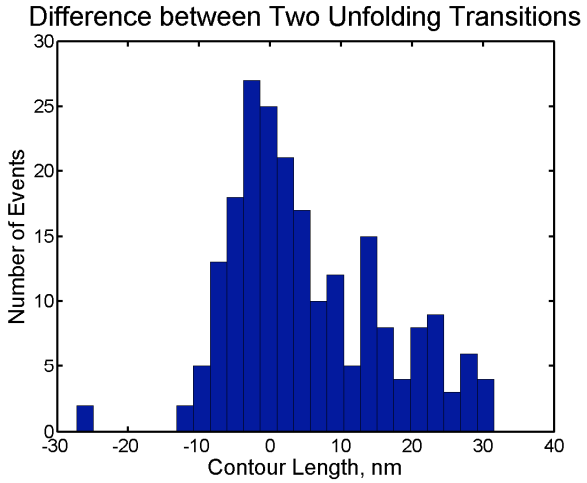
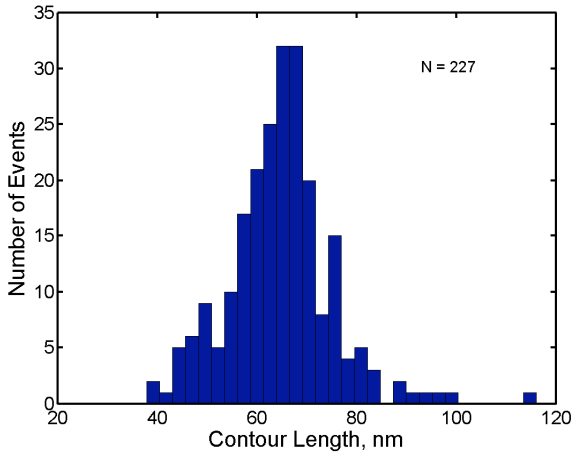
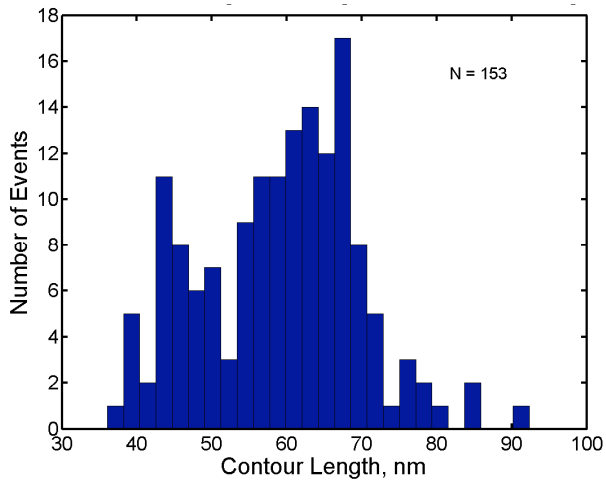


Figure 5.5: Histograms showing the (a) sum and (b) difference of the contour length changes of the unfolding transitions when two events are detected. The difference was calculated as ΔL_c(rip 2) – ΔL_c(rip 1).

(a)


N = 227
Number of Events
Contour Length, nm
0
5
10
15
20
25
30
35
20
40
60
80
100
120

(b)


N = 153
Number of Events
Contour Length, nm
0
2
4
6
8
10
12
14
16
18
30
40
50
60
70
80
90
100

(c)

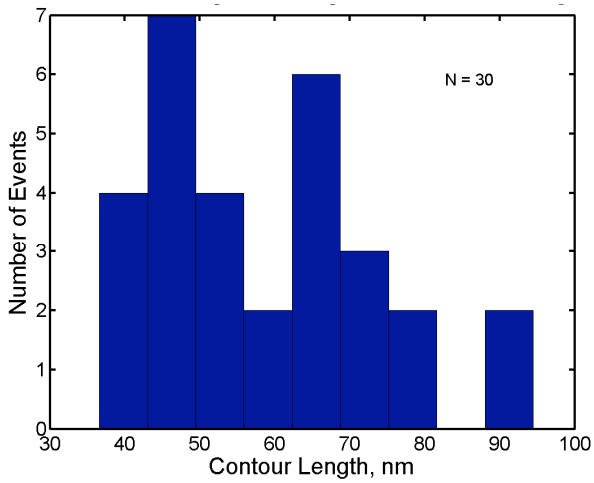

N = 30
Number of Events
Contour Length, nm
0
1
2
3
4
5
6
7
30
40
50
60
70
80
90
100

Figure 5.6: Histograms representing the total contour length change associated with a given force-extension curve after waiting (a) 5 seconds, (b) 3 seconds, and (c) 1 second at 1 pN. N is the total number of events represented.