Title
Detection and characterization of partially folded forms on the protein energy landscape

Author
Bernstein, Rachel Simma

Publication Date
2011

Peer reviewed|Thesis/dissertation

Detection and characterization of partially folded forms on the protein energy landscape

by

Rachel Simma Bernstein

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 P. Klinman, Co-chair
Professor David E. Wemmer
Professor Teresa Head-Gordon

Spring 2011

Abstract

Detection and characterization of partially folded forms on the protein energy landscape

by

Rachel Simma Bernstein

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Susan Marqusee, Co-chair
Professor Judith Klinman, Co-chair

Most proteins spend the majority of their time in their folded native state. Adopting this conformation, however, involves passing through various partially folded forms, including transition states and potentially kinetic intermediates. Furthermore, even under conditions favoring the folded conformation, a protein will take excursions away from the native state, populating partially and fully unfolded conformations. All of these states together constitute the protein energy landscape, and exploration of this complete landscape is crucial for a complete understanding of protein folding behavior, as well as potentially function. This work discusses a variety of methods applied to various model proteins, elucidating novel details about the folding landscapes.

Kinetic investigations were conducted with both T4 lysozyme and *E. coli* ribonuclease H (RNase H). The lysozyme study revealed a "hidden" unfolding intermediate in addition to the previously characterized folding intermediate, resolving a long-standing discrepancy between kinetic and native-state hydrogen exchange experiments. The RNase H study, on the other hand, focused on the nature of the transition state. This protein is known to fold through an intermediate, which has been investigated by various methods, including kinetic hydrogen exchange and kinetic analysis; however, the post-intermediate transition state had not previously been thoroughly investigated. The results from the study in this thesis suggest that the protein traverses the rate-limiting transition state through a highly localized nucleation-condensation process involving part of the protein that is unfolded in the kinetic intermediate.

E. coli RNase H was further investigated using a novel technique called native-state thiol alkyl-proton exchange (NSSX), a method analogous to native-state hydrogen exchange that takes advantage of the unique reactivity of cysteine to monitor exchange at the side chain, rather than the amide position of the backbone that is the target of hydrogen exchange experiments. Initial studies indicated that the wild-type protein was not amenable for these studies, but introduction of a stabilizing mutation allowed for investigation of the folding landscape under native conditions, with probes exposed on the folded and unfolded sides of the rate-limiting barrier exchanging in different kinetic regimes. This kinetic partitioning allowed for identification and characterization of novel partially folded species on the native side of the barrier and revealed

structural and kinetic data for probes that are only exposed on unfolded side of the barrier. Interestingly, some of the probes involved in the rate-limiting nucleation step, as identified in the kinetic analysis, are also shown to be structured in the transition state by the NSSX experiments.

These *in vitro* studies are complemented by *in vivo* translational misincorporation experiments with two pairs of homologous proteins. A library of mutant tRNAs was developed for NSSX to introduce cysteines in the place of a given amino acid during translation; however, it was found that *E. coli* RNase H was refractory to the misincorporation method. A highly similar protein from a thermophilic organism, *Thermus thermophilus*, on the other hand, shows robust misincorporation. Similarly, *E. coli* phosphoglycerate kinase (PGK) shows essentially no misincorporation, while yeast PGK misincorporates well. There is some *in vitro* evidence that those proteins that show significant misincorporate – *T. thermophilus* RNase H and yeast PGK – adopt partially folded conformations that are not accessible to their homologs. Therefore, it is plausible that misincorporation efficiency may report on the existence of partially folded forms *in vivo*; specifically, the absence of such conformations may result in degradation of the nascent chain on the ribosome, while adopting a protected conformation may allow for translation of the full-length misincorporated protein. While these results are preliminary and the hypothesis must be verified by further experiments, they provide an intriguing suggestion for a new *in vivo* probe of partially folded structure.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 6

ACKNOWLEDGEMENTS

I would never have made it through my graduate school career without the help of many, many people. First, I would like to thank my advisor, Susan Marqusee, for her tireless support and continued advocacy, both in my research and in my pursuits outside the lab. I would not be where I am today without her guidance.

I would also like to thank my parents, Joanne and Chuck, for their equally important non-scientific support. They have been by my side from the beginning, encouraging my intellectual endeavors no matter the direction or goal, and I can't thank them enough for everything they have done for me during my many years of schooling. I'm proud to say that it hasn't been all one-sided; over the years my dad has perfected a five minute elevator spiel explaining my research, just in time for my graduation.

Finally, I have to thank my many friends and coworkers who have seen me through the good times and the not-so-good times as I have worked toward my degree. There are too many people to name them all, but I would like to acknowledge my labmates Jesse Dill, Tracy Young, Katie Tripp, Sabriya Rosemond, Phillip Elms, Katelyn Connell, and Kath Ratcliff and my classmates Lee Bishop and Courtney Hastings, who provided endless scientific and personal encouragement. In addition, two people deserve special thanks. My classmate Katie Berry has been a great partner and confidante as we have made our way through our years at Berkeley, and my good friend Marissa Gunst has been a constant cheerleader.

This list is only the tip of the supportive iceberg I have relied on for the last six years. Without all of these people, this work would not have been possible; my deepest thanks go out to all of you.

Chapter 1

Introduction

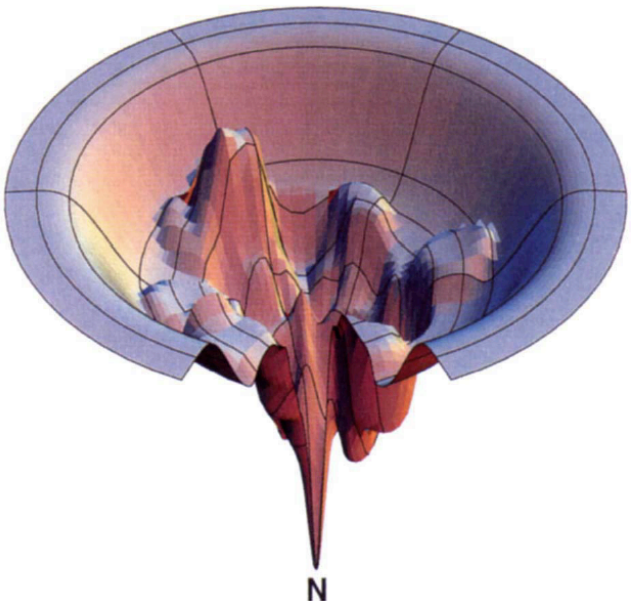
1.1 The protein folding problem

Proper protein function is vital for many cellular processes, and in almost all cases this function is determined by and dependent upon the protein's three-dimensional structure. Amazingly, this complex structure is determined by the relatively simple linear amino acid sequence (1). Adopting the native structure, however, requires that the linear polypeptide chain undergo a complex transformation from a disordered unfolded state to the functional folded conformation. If this process does not proceed properly, the resulting misfolded proteins can result in pathologies, including various neurodegenerative disorders and prion-based diseases. The vast majority of proteins, however, are able to fold properly and efficiently, in some cases taking less than milliseconds.

Therefore, it is clear that protein folding does not occur by a random search of all the conformational space available to a polypeptide chain; such a search would on average take billions of years or longer to find the native state, even for a relatively small protein. Thus, it is believed that the amino acid sequence codes not only for the protein's native, three-dimensional structure, but also for a specific pathway that allows folding to the structure to occur efficiently on a biologically relevant timescale (2).

Such pathways have been conceptualized with the folding funnel (3) (Figure 1), which describes all the conformations available to a given polypeptide chain. The rim of the funnel represents the many high-energy unfolded conformations; as the funnel narrows, the conformational space becomes more restricted and the energy of the conformations decreases down to the tip of the funnel, which represents the single lowest energy three-dimensional structure. The funnel also describes the kinetic barriers that must be crossed to reach lower-energy conformations; it is expected that a protein's preferred folding pathway will be described by the path down the funnel with best combination of low-energy species and low kinetic barriers.

Figure 1. A rendering of a theoretical protein folding funnel (3).



1.2 Partially folded species on the landscape

The ruggedness of this theoretical funnel reflects the wide variety of partially folded conformations available to any given polypeptide chain, including kinetic barriers (local maxima) and intermediates (local minima); a number of such species may be populated on a folding trajectory for any given protein. Therefore, investigation of such states can yield crucial insight into the nature of the folding funnel and the principles underlying efficient folding.

For example, there is growing evidence that transiently populating partially structured intermediates – which are being found in an increasing number of systems, including many that had been previously characterized as two-state (see, for example, (4-14)) – may aid efficient folding (13, 15-20). These results suggest that intermediates may play a crucial role in guiding the protein folding trajectory and narrowing the conformational search of the landscape to promote efficient folding. In addition, folded proteins may take excursions away from their native states to sample partially folded species, which may or may not correspond to kinetic intermediates on the unfolding trajectory. Furthermore, these partially folded forms are likely biologically relevant; for example, small motions away from the native state may be important for enzyme activity (21).

Both the barriers and intermediates are generally rare and only transiently populated, however, making them difficult to study. Despite these obstacles, various methods have been developed to investigate them, including Φ-value analysis and hydrogen exchange. Kinetic experiments, including Φ-value analysis (Section 1.3.2), have been a powerful tool to identify and structurally characterize the transition state for a number of proteins, including barnase (22), various SH3 domains (23-27), chymotrypsin inhibitor 2 (28), and others (29-40) and, in some cases, folding intermediates (see, for example, (4, 32, 36, 41)), but until recently, unfolding intermediates were much more difficult to characterize by such methods (11, 42). Hydrogen exchange (Section 1.3.3) is another valuable technique that has been exploited to reveal kinetic intermediates as well as partially unfolded species present at equilibrium (see, for example, (11, 43-48)). Unlike kinetic approaches, however, hydrogen exchange generally does not interrogate the rate-limiting transition state, focusing instead on marginally stable intermediates.

In the next section, I describe in more detail some common methods to detect and structurally characterize folding and unfolding intermediates, as well as a new method developed called thiol proton-alkyl exchange that has proven extremely valuable in the endeavor to characterize partially unfolded forms across the entire reaction coordinate.

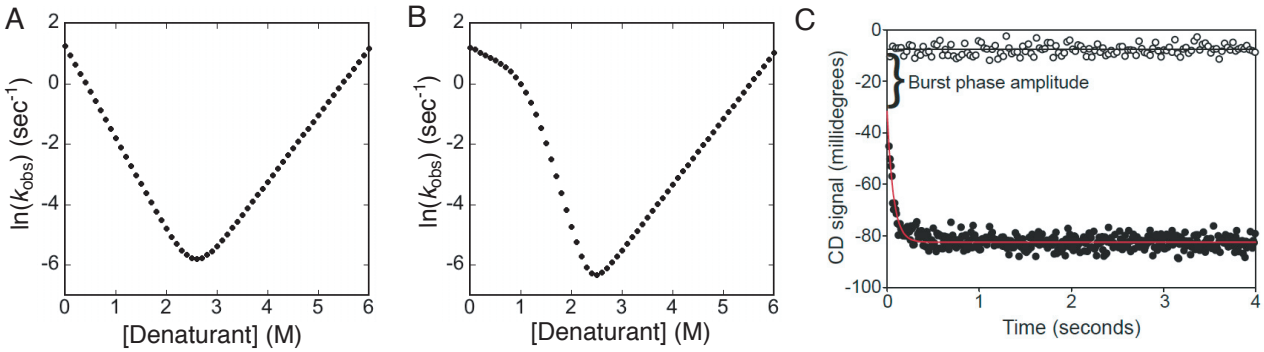
1.3 Experimental methods for exploring transition states and intermediates

1.3.1 Folding and unfolding kinetics: Chevron plots

While many proteins fold in an apparent two-state manner when evaluated at equilibrium, kinetic measurements often reveal the presence of transient intermediates. Such experiments usually involve the rapid dilution of folded protein into unfolding conditions, or vice versa, either with a stopped-flow instrument (dead time on the order of milliseconds) or rapid manual mixing (dead time on the order of seconds). The signal, often CD or fluorescence, can then be monitored as the protein either folds or unfolds. A plot of the natural log of the observed rate as a function of denaturant concentration results in what is commonly referred to as a chevron plot (Figure 2A, B). The minimum of the chevron occurs at the midpoint of the equilibrium unfolding transition

(C_m), and the left limb ([denaturant] below the C_m) is dominated by the folding rate, while the right limb ([denaturant] above the C_m) is dominated by the unfolding rate.

Figure 2. Model results from kinetic experiments. (A and B) Chevron plots, in the (A) absence and (B) presence of a folding intermediate formed within the dead time of the mixing. (C) Raw data demonstrating a burst phase amplitude formed in the dead time of the experiment.



In the absence of an intermediate, such analysis will result in observed rates that account for the total amplitude change and can be fit with a single exponential and a chevron plot with linear folding and unfolding limbs, resulting in the classic "V" shape (Figure 2A). Both limbs are linear and can be described by the following equations:

$$U \underset{k_{NU}}{\overset{k_{UN}}{\rightleftharpoons}} N$$

$\ln(k_{obs}) = \ln[k_{UN} + k_{NU}]$
$k_{UN}(den) = k_{UN}(H_20)\bullet \exp(-m^{\ddagger}_{UN}\bullet[den]/RT)$
$k_{NU}(den) = k_{NU}(H_20)\bullet \exp(-m^{\ddagger}_{NU}\bullet[den]/RT)$

Deviations from linearity are often seen, however, and may suggest the presence of a transient intermediate. Although there are a number of possible explanations for curvature (see Chapter 2), nonlinear chevron plots are often fit to an on-pathway, obligatory folding intermediate model (49):

$$U \underset{}{\overset{K_{UI}}{\rightleftharpoons}} I \underset{k_{ni}}{\overset{k_{in}}{\rightleftharpoons}} N$$

$\ln(k_{obs}) = \ln[f_I\bullet k_{IN} + k_{NI}]$
$f_I = K_{UI} / (1 + K_{UI})$
$k_{IN}(den) = k_{IN}(H_20)\bullet \exp(-m^{\ddagger}_{IN}\bullet[den]/RT)$
$k_{NI}(den) = k_{NI}(H_20)\bullet \exp(-m^{\ddagger}_{NI}\bullet[den]/RT)$
$K_{UI}(den) = K_{UI}(H_20)\bullet \exp(-m_{UI}\bullet[den]/RT)$

The presence of an intermediate can also be inferred by an observable burst phase amplitude, in which the extrapolated signal at zero time is different from the initial signal expected for the fully unfolded conditions (Figure 2C). This signal change is attributed to the rapid acquisition of structure during the dead time of the experiment that occurs in unobserved fast phase(s). If this fast phase is much faster than the observed phase, this burst phase amplitude can be used to determine the stability of the intermediate (K_{UI}).

1.3.2 Φ-values

While simple kinetic studies can identify the existence of an intermediate, they do not yield information about its structure or the structure of the transition state. Introducing single-site mutants and observing the effect on the measured kinetics relative to the equilibrium stability, called Φ-value analysis, however, can begin to reveal the side chains involved in stabilizing these high-energy species (22, 41). For such experiments, destabilizing point mutations are introduced, and the effect of the mutation on the equilibrium stability and folding and unfolding rates are measured. Simply, the Φ-values are the ratio of the change in stability of an intermediate or the transition state to the overall change in stability:

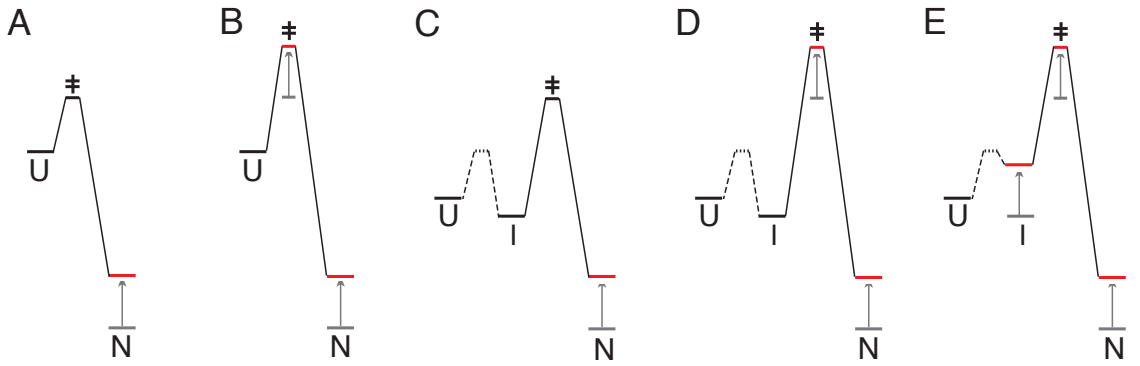
$$\Phi^{\ddagger} = \Delta\Delta G_{U\ddagger}/\Delta\Delta G_{UN} \qquad \Phi^{I} = \Delta\Delta G_{UI}/\Delta\Delta G_{UN}$$

From this information, it can be inferred whether a given position is involved in energetic interactions in either the transition state or an intermediate, beginning to create a structural picture for each of these species.

Consider, for example, a simple, two-state protein with no detectable intermediates. If a mutation is introduced that destabilizes the native state (as determined by equilibrium denaturant melt) but not the transition state (as determined by the unfolding rate), $\Delta\Delta G_{U\ddagger}$ will be 0, leaving the folding rate unchanged and resulting in a $\Phi^{\ddagger}$-value of 0 (Figure 3A). The general interpretation for a $\Phi^{\ddagger}$-value close to 0, therefore, is that the side chain at that position is not involved in structured energetic interactions in the transition state. On the other hand, if a mutation destabilizes the native state and the transition state to the same degree, both $\Delta\Delta G_{U\ddagger}$ and $\Delta\Delta G_{UN}$ will be equal and non-zero, leaving the unfolding rate unchanged and resulting in a $\Phi^{\ddagger}$-value of 1 (Figure 3B). This $\Phi^{\ddagger}$-value can thus be interpreted to mean that the position being probed has already adopted native-like energetic interactions in the transition state. This protein engineering analysis, however, makes the simple assumption that the mutations perturb only the energetics and not the structure of the protein.

In a case where there is a populated intermediate with measurable stability, the same logic can be applied to measure both $\Phi^{\ddagger}$ and Φ^{I}, which reflects the interactions present in the intermediate. If a mutation destabilizes the native state but not the transition state or the intermediate, the folding rate and intermediate stability will be unchanged ($\Delta\Delta G_{U\ddagger} = \Delta\Delta G_{UI} = 0$) and both $\Phi^{\ddagger}$ and Φ^{I} will be 0 (Figure 3C), suggesting that the side chain does not contribute to the stability of either the transition state or the intermediate. If a mutation destabilizes the native state and the transition state to the same degree, but leaves the intermediate unaffected ($\Delta\Delta G_{U\ddagger} = \Delta\Delta G_{UN}$; $\Delta\Delta G_{UI} = 0$), the unfolding rate will be unchanged, resulting in a $\Phi^{\ddagger}$ of 1 but a Φ^{I} of 0 (Figure 3C). Finally, if a mutation equally destabilizes the native state, transition state, and intermediate ($\Delta\Delta G_{UN} = \Delta\Delta G_{U\ddagger} = \Delta\Delta G_{UI}$), both $\Phi^{\ddagger}$ and Φ^{I} will be 1 (Figure 3E).

Figure 3. Model reaction coordinates illustrating results from Φ-value experiments. A and B show possibilities for Φ-values in proteins without intermediates; C, D, and E show possibilities for Φ-values in proteins with detectable kinetic folding intermediates, with the barrier between U and I shown in dotted lines to represent that it is generally traversed within the dead time of the experiment and thus not measured. (A) A variant of a two-state protein that destabilizes the native state but not the transition state will speed unfolding but leave the folding rate unchanged, resulting in a Φ-value of 0 ($\Delta\Delta G_{U\ddagger} = 0$). (B) A variant of a two-state protein that destabilizes both the native state and the transition state will slow folding but leave the unfolding rate unchanged, resulting in a Φ-value of 1 ($\Delta\Delta G_{U\ddagger} = \Delta\Delta G_{UN}$). (C) A variant of a three-state protein that destabilizes the native state but not the transition state or the intermediate will speed unfolding but leave both the folding rate and the stability of the intermediate unchanged, resulting in both a Φ^I and $\Phi^\ddagger$ of 0 ($\Delta\Delta G_{U\ddagger} = \Delta\Delta G_{UI} = 0$). (D) A variant of a three-state protein that destabilizes the native state and the transition state but not the intermediate will slow folding but leave the unfolding rate and stability of the intermediate unchanged, resulting in $\Phi^\ddagger$ of 1 a Φ^I of 0 ($\Delta\Delta G_{U\ddagger} = \Delta\Delta G_{UN}$; $\Delta\Delta G_{UI} = 0$). (E) A variant of a three-state protein that destabilizes the native state, transition state, and intermediate will slow folding but leave the unfolding rate intermediate unchanged, resulting in both a $\Phi^\ddagger$ and Φ^I of 1 ($\Delta\Delta G_{UI} = \Delta\Delta G_{U\ddagger} = \Delta\Delta G_{UN}$).



Φ-values of 0 or 1 are relatively easy to interpret based on this model, but kinetic experiments can also yield fractional or non-canonical Φ-values (<0 or >1), which are more difficult to rationalize. Fractional Φ-values arise when the intermediate or transition state is destabilized, but not to the extent of the native state. Such behavior could result from a number of causes; for example, a Φ-value of 0.5 could indicate that the side chain has a partial energetic contribution to the state in question, or that there is an ensemble of structures in which the energetic interactions at the position are fully formed in half the conformations and completely unformed in the other half (50, 51).

Φ-values greater than 1 or less than 0 are often attributed to non-native interactions in the transition state (52). An alternate explanation comes from a computational study suggesting that non-classical Φ-values could be attributed to parallel, coupled folding pathways rather than a one-dimensional folding reaction coordinate – as suggested by the commonly referenced folding funnel (53). Others argue, however, that such values are simply due to inadequate destabilization, resulting in unreliable analysis (54, 55). Because Φ-values are calculated as a

ratio of free energy differences, with $\Delta\Delta G_{NU}$ in the denominator, this number must be reliable and large enough that the resulting Φ-value is not overly affected by small deviations in the value. The exact destabilization appropriate for such analysis, however, is still a matter of debate (51, 54, 56-58).

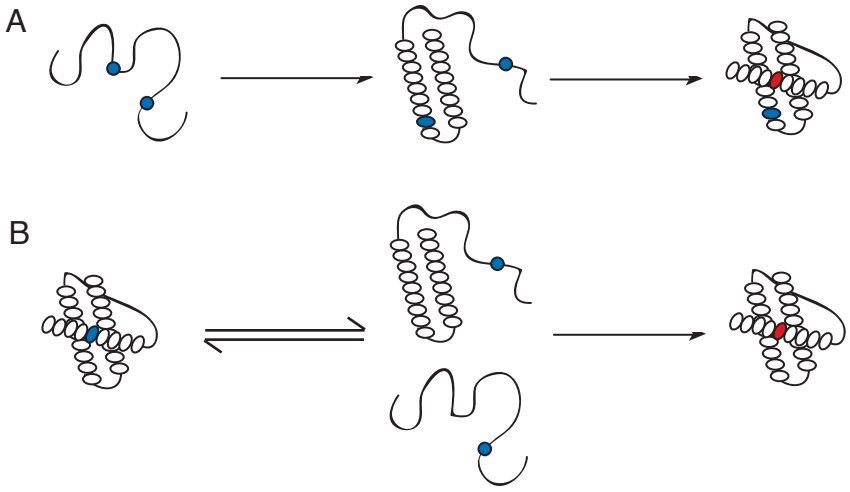
Despite these limitations, Φ-value analysis has been very successful at elucidating the energetic contributions of intermediates and transition states in a variety of model systems. It has been used to infer the transition state structure for a number of proteins, including barnase (22), various SH3 domains (23-27, 40), chymotrypsin inhibitor 2 (28), and others (29-39). These studies reveal large amounts of variation between transition states for different proteins; the suggested structures range from highly polarized (24, 34, 35), with some interactions essentially fully formed and others absent, to states with more diffuse structure (29, 36), where many interactions are partially formed.

Most of these studies have focused on two-state proteins and thus yielded $\Phi^{\ddagger}$, allowing the interrogation of barriers along the reaction coordinate and inference of structures of high-energy transition states. In the case of 3-state proteins in which the Φ^{I} can be investigated, this method can also provide some insight into the structure of stable intermediates along the folding pathway. For example, Φ-value studies of barnase (22, 41), barstar (39), the Fyn SH3 domain (40), and Im7 (36), all of which fold via kinetic intermediates, have measured both $\Phi^{\ddagger}$ and Φ^{I}. In general, however, the strength of Φ-value analysis is to investigate the regions involved in the rate-limiting transition states rather than intermediates.

1.3.3 Hydrogen exchange

Hydrogen exchange is a powerful and complementary method that has been widely used to detect and characterize partially folded intermediates present on the energy landscape (see (59) for a review). In such experiments, exchange of backbone amide protons with solvent deuterons (or vice versa) is monitored, usually by NMR or mass spectrometry. Generally, protons that are buried and/or hydrogen-bonded will exchange slowly, while protons exposed to the solvent and/or not involved in hydrogen bonds will exchange more rapidly. Thus, such experiments can reveal structural information about the solvent exposure and backbone hydrogen bonding of each residue, and all probes that exchange on a time-scale compatible with the monitoring method can be characterized simultaneously, providing a relatively complete structural picture of the species being explored. Furthermore, these experiments can be carried out to detect either the kinetics of a conformational change or the equilibrium distribution of conformations to, for example, reveal the structure of an early folding intermediate by kinetic hydrogen exchange (see, for example, (60-69)), or detect rare partially unfolded forms present at equilibrium by native-state hydrogen exchange (see, for example, (43-46, 70-73)).

Figure 4. Schematic models of (A) kinetic labeling experiment and (B) native-state exchange experiments. (A) Kinetic experiments are initiated by the rapid dilution of unfolded protein into folding conditions such that probes that become protected early in the folding trajectory do not exchange (blue), while those that become structured later in the trajectory do exchange (red). (B) Native-state experiments are conducted under native conditions; exchange results from fluctuations away from the native-state to various "open" forms (either partially or fully unfolded).



Kinetic labeling experiments are usually conducted either by competition or pulse labeling. In competition experiments, unfolded deuterated (or protonated) protein is rapidly diluted into protonated (or deuterated) refolding conditions. Observed exchange rates at each position depend on the relative rates of exchange versus formation of a protected state. Specifically, positions that do not exchange can be assumed to be part of a protected structure early in the folding trajectory that forms more quickly than the intrinsic exchange rate from the unfolded state; those that do exchange remain unprotected long enough for exchange to occur. Therefore, these experiments reveal probes that gain protection early in the folding trajectory, i.e. folding intermediates. Such experiments have been carried out with lysozyme from hen egg white (74) and T4 (75), for example, revealing structural information about folding intermediates.

Furthermore, these experiments can be conducted using refolding conditions at varying pH, thus varying the intrinsic exchange rate for the fully unfolded state. In such experiments, the proton occupancy is 50% at the pH where the folding rates are equal to the intrinsic exchange rates, which have been determined using model peptides (76). Therefore, these experiments provide not only structural information, but also yield a rate for formation of protection at each probe position. Such hydrogen exchange competition experiments have been successfully conducted with ribonuclease A (77), ribonuclease H (78), and others (79, 80).

Similar to these competition experiments, pulse labeling hydrogen exchange also involves rapidly initiating refolding of deuterated protein into protonated buffer, or vice versa; however, instead of direct competition between exchange and folding, a pulse is applied to initiate exchange at specific time points, often accomplished by a change in pH. These experiments are used to construct a time course for formation of probe protection along the refolding trajectory,

providing structural information about intermediates on the landscape for proteins such as ribonuclease A (61, 81), cytochrome c (60), barnase (62, 64), T4 lysozyme (63), ribonuclease H (69), and others (65-68). Importantly, however, such experiments can only detect species that are populated before the rate-limiting transition state.

Native-state hydrogen exchange experiments, on the other hand, have the power to access the entire energy landscape under equilibrium native conditions. Though the native state is still the predominantly populated species, rare conformations are also populated according to their Boltzmann distribution. These species can be interrogated based on the Linderstrom-Lang model for the exchange process(82). According to this model, an amide hydrogen in the "closed" or native conformation must undergo a fluctuation to some alternative "open" conformation in order to exchange:

$$closed \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} open \xrightarrow{k_{int}} exchanged \quad \text{(scheme 1)}$$

The detection method is blind to the closed conformation, allowing detection of the rare open species. This method can therefore detect and characterize very small population of high-energy exchanging conformations populated at equilibrium under native conditions.

The information revealed about these partially unfolded species depends upon the kinetic regime of exchange. In scheme 1 above, k_{op} and k_{cl} refer to the rates of opening to and closing from that exchange-competent conformation, while k_{int} refers to the intrinsic rate of exchange. Based on this model, the observed rate of exchange for any probe can be written as:

$$k_{obs} = k_{op} \bullet k_{int}/(k_{cl} + k_{int})$$

This expression can be simplified in two extreme regimes. In the EX1 regime, where the closing rate is much slower than the intrinsic rate of exchange ($k_{cl} << k_{int}$):

$$k_{obs} = k_{op} \quad \text{(EX1)}$$

Such experiments thus directly reveal kinetic information about the rates of opening to various exchange-competent species. In the EX2 regime, where closing is much faster than intrinsic exchange ($k_{cl} >> k_{int}$):

$$k_{obs} = k_{int} \bullet k_{op}/k_{cl} = k_{int} \bullet K_{op} \quad \text{(EX2)}$$

In this case, stabilities can be calculated:

$$\Delta G = - RT \ln(K_{op}) = - RT \ln(k_{obs}/k_{int})$$

Most NSHX experiments are conducted in the EX2 regime, and therefore reveal information about the stabilities of different species populated at equilibrium. EX2 NSHX experiments have successfully detected high-energy partially unfolded conformations in a number of proteins, including cytochrome c (43, 46), ribonuclease H (44), T4 lysozyme (83), and others (45, 71, 72, 84); the lack of kinetic information from these experiments, however, means that these species cannot be positioned on the folding landscape.

Other studies have used a combination of EX1 and EX2 NSHX data to localize intermediates on the reaction coordinate in addition to structural characterization, for example in turkey ovomucoid third domain (85), staphylococcal nuclease (48), ubiquitin (86), and others (70, 73, 87-90). There has also been some work toward using NSHX to access the EX1 regime exclusively (91), for example with ubiquitin (92) and the SH3 domain (93), but such approaches remain relatively undeveloped.

Despite their utility, hydrogen exchange experiments are subject to significant limitations. As mentioned previously, they have been used most successfully to probe intermediates on the landscape, but the traditional approaches do not reveal significant information about the barriers separating these conformations. Furthermore, exchange reports only on the conformation and hydrogen-bonding network of the backbone and is thus blind to conformational fluctuations involving only the side chains. Finally, the site-specific nature of hydrogen exchange experiments relies heavily on NMR; therefore, proteins that are not amenable to NMR analysis, such as large or poorly behaved proteins, are generally poor for hydrogen exchange experiments.

1.3.4 Thiol proton-alkyl exchange

The limitations to traditional hydrogen exchange methods illustrate the need for continued development of new techniques. In particular, thiol exchange has emerged in the past few years as a potential complement to hydrogen exchange. The same formalisms are applied to both techniques, but instead of monitoring exchange of the backbone amide proton, thiol exchange experiments take advantage of the unique reactivity of the cysteine side chain and monitor its reactivity with modifying reagents. Analogous to hydrogen exchange, thiol exchange can also be applied in either kinetic or equilibrium experiments to detect kinetic intermediates as well as partially unfolded equilibrium species.

There are, however, some distinct differences between the two approaches beyond probing exchange at backbone positions versus side chain. Silverman and Harbury have devised a simple, gel-based assay to simultaneously characterize the exchange behavior of a number of cysteine probes in a single experiment (94), eliminating the need for NMR or mass spec to determine the site of modification. Therefore, such experiments can be conducted with proteins that are too large or poorly behaved for NMR. Furthermore, the potential to use a number of different cysteine modifying reagents opens up many possibilities not available to hydrogen exchange experiments, including, for example, varying the intrinsic modification rate to access different exchange regimes, or varying the size of the modifying agent to investigate the magnitude of the structural openings that allow exchange to occur.

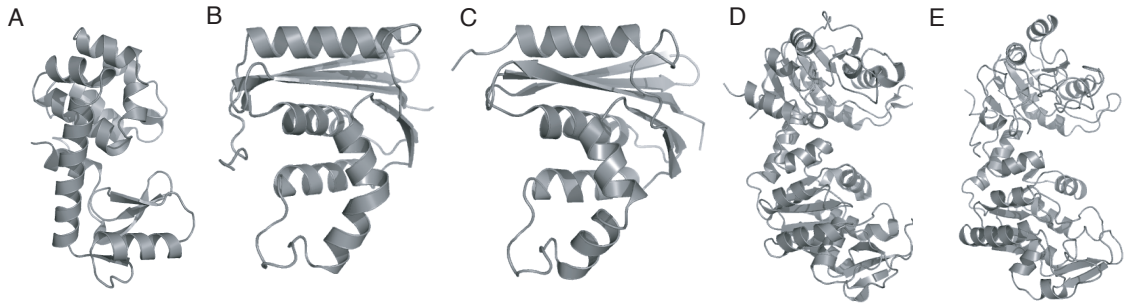
Of course, the primary limitation for such experiments is the need for cysteine side chains at the sites where one would like to monitor exchange. Cysteine side chains can easily be introduced by site-directed mutagenesis, but these non-native cysteines may perturb the protein's energy landscape. Even if slight perturbations are introduced, however, such an approach can still reveal much novel information about a protein's folding landscape. Thus, while it is important to keep in mind the potential destabilizing effects of such mutations, with careful controls and analysis they need not invalidate the results of this approach.

Thiol exchange investigations of various types have already been pursued to monitor the energetics of conformational changes such as unfolding in a number of systems (95-102). One of the most comprehensive studies characterized the native-state thiol proton-alkyl exchange behavior of the yeast TIM protein, revealing two partially unfolded species present at equilibrium (103). This protein was chosen because it is not amenable to most other traditional biophysical characterization methods, proving the utility of the thiol exchange approach; however, the lack of previous characterization made it difficult to fully appreciate the strengths and limitations of this new method. Therefore, one of the major goals in pursuing the development of this method was to apply it to a well-characterized protein so there would be a solid foundation of knowledge upon which to interpret the thiol exchange results.

1.4 Model systems

The work in this thesis spans experiments on a number of model systems. Crystal structures for each of these proteins are shown in Figure 5, and they are described in some detail in the following section.

Figure 5. Structural representations based on X-ray crystallography for all model proteins discussed in this thesis. (A) T4 lysozyme. (B) *E. coli* ribonuclease H. (C) *T. thermophilus* ribonuclease H. (D) *E. coli* phosphoglycerate kinase. (E) Yeast phosphoglycerate kinase.



1.4.1 Bacteriophage T4 lysozyme

T4 lysozyme is a 164-residue, mainly α-helical protein that cleaves the glycosidic bond in some polysaccharides, including those in bacterial cell walls, resulting in cell lysis. It is composed of two sequentially discontinuous subdomains, the N-terminal subdomain and the C-terminal subdomain; the helix at the N-terminus of the protein is structurally part of the C-terminal subdomain. By equilibrium denaturation, the protein behaves in a two-state manner, but hydrogen exchange studies identify two different intermediates: pulse-labeling hydrogen exchange detects an early folding intermediate with portions of both subdomains protected from exchange (63), while native-state hydrogen exchange reveals a high-energy partially unfolded form with structure in just the C-terminal subdomain(83). The results described in this thesis suggest that the protein actually folds through two kinetic intermediates, one on each side of the rate-limiting barrier for folding, and that the different HX experiments are each biased to observe one or the other.

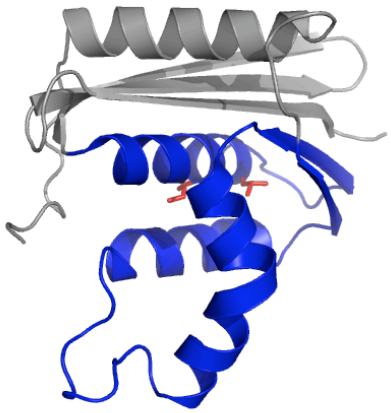
1.4.2 Ribonuclease HI (RNase H)

RNase H is a small (155 amino acid in the case of the *E. coli* protein), single-domain α+β protein that cleaves the RNA strand of DNA/RNA hybrids. This protein is highly conserved

across many bacteria; in this thesis I will focus on work done with the *E. coli* protein, and a small amount of work done with the protein from *Thermus thermophilus*.

Equilibrium denaturation studies of *E. coli* RNase H are adequately fit with a two-state model, but kinetic experiments have revealed that it folds through a kinetic folding intermediate, as shown both by rollover in the folding limb and a substantial burst phase amplitude (104). Kinetic hydrogen exchange experiments reveal that the core helices are protected within this burst phase (within 10 msec of initiating folding) (69), and Φ-value analysis also indicates that these core helices are structured in the intermediate, while the so-called periphery (strands 1-3 and helix E) is not (32, 105) (Figure 5). Native-state hydrogen exchange experiments also detect a high energy species similar to the kinetic intermediate, with the core helices folded and the β sheet and helix E unfolded(44). Single-molecule mechanical studies reveal an obligatory folding intermediate; the measured extension (end-to-end distance) for this species agrees with the model suggested by the hydrogen exchange and Φ-value results, and mutations have a similar affect on this mechanical intermediate (106). Together, these results support a model in which the protein undergoes an early collapse of the core helices into a marginally stable folding intermediate that is briefly populated until the protein is able to traverse the rate-limiting barrier to folding. Furthermore, Φ-value analysis has been used to show that the folding trajectory is robust to even extreme changes in the distribution of stability throughout the protein (105).

Figure 6. A simple structural model of the *E. coli* RNase H intermediate. The core region, believed to be structured in the intermediate, is colored blue; the periphery region, believed to be unstructured in the intermediate, is shown in grey. The two side chains thought to be involved in residual structure in the unfolded states of *T. thermophilus* and *C. tepidum* RNases H, I53 and L56, are shown in red sticks.



T. thermophilus RNase H is 52% identical in protein sequence to *E. coli* RNase H and shares a very similar native structure as determined by X-ray crystallography (107) (Figure 5B, C). The protein from *T. thermophilus* has a significantly higher melting temperature than that from *E. coli* (86°C and 67°C, respectively) and is more stable at all temperatures (108), but both proteins appear to fold through structurally similar kinetic intermediates (109); like for the *E. coli* protein, native-state hydrogen exchange for the *T. thermophilus* protein also reveals a conformation similar to that seen in the kinetic experiment (110).

There are some behaviors that are unique to the *T. thermophilus* protein, however. The native-state hydrogen exchange experiments also detected a second, novel species with just helix E unfolded (110), and the folding kinetics have an additional slow phase that is not seen for the *E. coli* protein and is likely related to proline isomerization (109). Finally, the *T. thermophilus* protein has a lower ΔC_p than the *E. coli* protein, which is attributed to residual structure in the unfolded state; this structure may be related to the ability of the *T. thermophilus* protein to function at temperatures at which the *E. coli* protein is unfolded(108). Work with chimeric proteins and *Chlorobium tepidum* RNase H, a protein from a moderately thermophilic organism that has a melting temperature similar to that for *E. coli* RNase H (66.5°C) but a ΔC_p similar to that of *T. thermophilus* RNase H (111), suggests that this residual structure occurs in the core region of the protein (111-113) (Figure 6). The exact structured region and its role on the energy landscape, however, remain to be elucidated.

1.4.3 Phosphoglycerate kinase (PGK)

PGK is a small, highly conserved glycolytic protein composed of two domains, each with a Rossmann fold topology, connected by an α-helical linker. The studies presented here focus on two homologs, *E. coli* PGK and yeast PGK, which are 39% identical and 56% similar in amino acid sequence. The three-dimensional structures as determined by X-ray crystallography are nearly identical (Figure 4D, E). Their thermodynamic stabilities are also similar ($\Delta G(H_20)$ = 9.3 and 9.7 kcal/mol, respectively (114)), but *E. coli* PGK is resistant to proteolysis while yeast PGK is not. Further studies showed that this difference arises from differing folding behavior of the domains: the isolated C-terminal domain of *E. coli* PGK cannot fold, while both domains of yeast PGK fold independently (114).

1.5 Summary of work in this thesis

In Chapter 2, I describe a study of the folding and unfolding kinetics of T4 lysozyme that revealed a "hidden intermediate" on the native side of the rate-limiting barrier. This study reconciled conflicting equilibrium, kinetic, and structural data by suggesting that two intermediates exist, one on each side of the rate-limiting barrier, and uses single-site mutants to infer the structured regions of these two species. This work was carried out in collaboration with Jason Celliti and published in the *Journal of Molecular Biology*; I am the second author.

In Chapter 3, I use a kinetic Φ-value analysis to investigate the transition state for the well-studied *E. coli* RNase H. This analysis suggests a crucial hydrophobic nucleation-condensation process after formation of the transition state. Unlike early events in folding that follow the nucleation-condensation mechanism, however, there is evidence that specific steric packing interactions, as well as the traditional hydrophobic collapse, is crucial for this later, post-intermediate transition state. This work is being prepared for submission; I will be the first author.

In Chapter 4, I continue to explore partially folded species on the landscape by describing my efforts to apply newly developed native-state thiol alkyl-proton exchange (NSSX) methodology to *E. coli* RNase H. These studies were complicated by the kinetic exchange regime: it was found that the wild-type protein did not exchange in either extreme, simplifiable regime, but instead exchanged predominately in the so-called EXX regime, where observed exchange rates are thus difficult to interpret. To bias exchange toward the EX1 regime, I then apply the NSSX to the

well-characterized variant D10A *E. coli* RNase H, which has greatly slowed refolding. I found that this variant could be used as a background to measure the exchange rates of cysteine probes placed throughout the protein, and that some probes exchange in EX1 while others exchange in EX2. This mixed exchange creates the possibility for an EX1/EX2 experiment that could simultaneously provide kinetic and equilibrium information to reveal detailed information about the ruggedness of the landscape.

In Chapter 5, I describe the results of the full NSSX characterization of D10A *E. coli* RNase H. The mixture of EX1 and EX2 exchange seen previously is shown to correlate to the exchanging species' position on the reaction coordinate; simply, EX2 probes exchange on the native side of the barrier, while EX1 probes exchange on the unfolded side of the barrier (although there are many possible explanation to this simple kinetic partitioning model, which are discussed). This mixed EX1/EX2 approach thus allows me to localize intermediates on the reaction coordinate in a way that is not possible in a traditional all-EX2 (i.e. traditional NSHX) experiment, and potentially allows detection of additional intermediates that were previously masked by lower energy intermediates involving exchange of the same probes. Furthermore, using the cysteine side chain as a probe instead of the backbone reveals some novel partially unfolded forms that were not seen in previous NSHX experiments. This work is currently under review in *Proceedings of the National Academy of Sciences*; I am the co-first author.

In Chapter 6, I move toward questions about *in vivo* folding behavior. Using misincorporator tRNAs developed by the Harbury lab – cysteine tRNAs with a mutated anticodon such that they introduce cysteine in the place of a desired amino acid – I observed differential behavior for two pairs of homologous proteins. While no misincorporation is seen in *E. coli* RNase H, *Thermus thermophilus* RNase H was found to misincorporate quite well. PGKs from *E. coli* and yeast were also tested; again, the *E. coli* protein did not misincorporate well, while the yeast protein did. The reason for these discrepancies remain unresolved, but I argue that could be related to differential folding behavior of the nascent chains while on the ribosome, resulting in degradation of nascent chains lacking structure but not those form some partially structured conformations during translation.

1.6 References

1. Anfinsen CB & Haber E (1961) Studies on the reduction and re-formation of protein disulfide bonds. *J Biol Chem* 236:1361-1363.
2. Levinthal C (1968) Are there pathways for protein folding? *J Chim Phys* 65(1):44-45.
3. Dill KA & Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4(1):10-19.
4. Matouschek A, Kellis JT, Jr., Serrano L, Bycroft M, & Fersht AR (1990) Transient folding intermediates characterized by protein engineering. *Nature* 346(6283):440-445.
5. Bachmann A & Kiefhaber T (2001) Apparent two-state tendamistat folding is a sequential process along a defined route. *J Mol Biol* 306(2):375-386.
6. Sanchez IE & Kiefhaber T (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* 325(2):367-376.
7. Teilum K, Maki K, Kragelund BB, Poulsen FM, & 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(15):9807-9812.
8. Ferguson N, Capaldi AP, James R, Kleanthous C, & Radford SE (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* 286(5):1597-1608.
9. Jemth P*, et al*. (2004) Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc Natl Acad Sci U S A* 101(17):6450-6455.
10. 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(4):757-764.
11. Feng H, Vu ND, & Bai Y (2005) Detection of a hidden folding intermediate of the third domain of PDZ. *J Mol Biol* 346(1):345-353.
12. Korzhnev DM*, et al*. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430(6999):586-590.
13. 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(46):16650-16655.
14. Khorasanizadeh S, Peters ID, Butt TR, & Roder H (1993) Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* 32(27):7054-7063.
15. Khorasanizadeh S, Peters ID, & Roder H (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* 3(2):193-205.
16. Heidary DK, O'Neill JC, Jr., Roy M, & Jennings PA (2000) An essential intermediate in the folding of dihydrofolate reductase. *Proc Natl Acad Sci U S A* 97(11):5866-5870.
17. Wagner C & Kiefhaber T (1999) Intermediates can accelerate protein folding. *Proc Natl Acad Sci U S A* 96(12):6716-6721.
18. Friel CT, Beddard GS, & Radford SE (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(1):261-273.
19. Goldbeck RA, Chen E, & Kliger DS (2009) Early events, kinetic intermediates and the mechanism of protein folding in cytochrome C. *Int J Mol Sci* 10(4):1476-1499.

20. Spudich GM, Miller EJ, & 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(2):609-618.
21. Henzler-Wildman KA, *et al*. (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450(7171):838-844.
22. Serrano L, Matouschek A, & Fersht AR (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(3):805-818.
23. Martinez JC, Pisabarro MT, & Serrano L (1998) Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol* 5(8):721-729.
24. Grantcharova VP, Riddle DS, Santiago JV, & Baker D (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* 5(8):714-720.
25. Fulton KF, Main ER, Daggett V, & Jackson SE (1999) Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J Mol Biol* 291(2):445-461.
26. Northey JG, Di Nardo AA, & Davidson AR (2002) Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol* 9(2):126-130.
27. Zarrine-Afsar A, Lin SL, & Neudecker P (2010) Mutational investigation of protein folding transition states by Phi-value analysis and beyond: lessons from SH3 domain folding. *Biochem Cell Biol* 88(2):231-238.
28. Itzhaki LS, Otzen DE, & Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254(2):260-288.
29. Bueno M, Ayuso-Tejedor S, & Sancho J (2006) Do proteins with similar folds have similar transition state structures? A diffuse transition state of the 169 residue apoflavodoxin. *J Mol Biol* 359(3):813-824.
30. Curnow P & Booth PJ (2009) The transition state for integral membrane protein folding. *Proc Natl Acad Sci U S A* 106(3):773-778.
31. Huysmans GH, Baldwin SA, Brockwell DJ, & Radford SE (2010) The transition state for folding of an outer membrane protein. *Proc Natl Acad Sci U S A* 107(9):4099-4104.
32. Raschke TM, Kho J, & Marqusee S (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* 6(9):825-831.
33. Calosci N, *et al*. (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc Natl Acad Sci U S A* 105(49):19241-19246.
34. Went HM & Jackson SE (2005) Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* 18(5):229-237.
35. Garcia-Mira MM, Boehringer D, & Schmid FX (2004) The folding transition state of the cold shock protein is strongly polarized. *J Mol Biol* 339(3):555-569.
36. Capaldi AP, Kleanthous C, & Radford SE (2002) Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* 9(3):209-216.
37. Bartlett AI & Radford SE (2010) Desolvation and development of specific hydrophobic core packing during Im7 folding. *J Mol Biol* 396(5):1329-1345.
38. Anil B, Sato S, Cho JH, & Raleigh DP (2005) Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing. *J Mol Biol* 354(3):693-705.

39. Nolting B, *et al*. (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci U S A* 94(3):826-830.
40. Neudecker P, Zarrine-Afsar A, Davidson AR, & Kay LE (2007) Phi-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc Natl Acad Sci U S A* 104(40):15717-15722.
41. Matouschek A, Serrano L, & Fersht AR (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(3):819-835.
42. Cellitti J, Bernstein R, & Marqusee S (2007) Exploring subdomain cooperativity in T4 lysozyme II: uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Sci* 16(5):852-862.
43. Bai Y, Sosnick TR, Mayne L, & Englander SW (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* 269(5221):192-197.
44. Chamberlain AK, Handel TM, & Marqusee S (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* 3(9):782-787.
45. Fuentes EJ & Wand AJ (1998) Local dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* 37(11):3687-3698.
46. Maity H, Maity M, Krishna MM, Mayne L, & Englander SW (2005) Protein folding: the stepwise assembly of foldon units. *Proc Natl Acad Sci U S A* 102(13):4741-4746.
47. Bai Y (2006) Protein folding pathways studied by pulsed- and native-state hydrogen exchange. *Chem Rev* 106(5):1757-1768.
48. Bedard S, Mayne LC, Peterson RW, Wand AJ, & Englander SW (2008) The foldon substructure of staphylococcal nuclease. *J Mol Biol* 376(4):1142-1154.
49. Baldwin RL (1996) On-pathway versus off-pathway folding intermediates. *Fold Des* 1(1):R1-8.
50. Fersht AR, Itzhaki LS, elMasry NF, Matthews JM, & Otzen DE (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc Natl Acad Sci U S A* 91(22):10426-10429.
51. Raleigh DP & Plaxco KW (2005) The protein folding transition state: what are phi-values really telling us? *Protein Pept Lett* 12(2):117-122.
52. Royer CA (2008) The nature of the transition state ensemble and the mechanisms of protein folding: a review. *Arch Biochem Biophys* 469(1):34-45.
53. Ozkan SB, Bahar I, & Dill KA (2001) Transition states and the meaning of Phi-values in protein folding kinetics. *Nat Struct Biol* 8(9):765-769.
54. Sanchez IE & Kiefhaber T (2003) Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J Mol Biol* 334(5):1077-1085.
55. de los Rios MA, Daneshi M, & Plaxco KW (2005) Experimental investigation of the frequency and substitution dependence of negative phi-values in two-state proteins. *Biochemistry* 44(36):12160-12167.
56. Fersht AR & Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci U S A* 101(21):7976-7981.
57. de los Rios MA, *et al*. (2006) On the precision of experimentally determined protein folding rates and phi-values. *Protein Sci* 15(3):553-563.
58. Naganathan AN & Munoz V (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proc Natl Acad Sci U S A* 107(19):8611-8616.

59. Englander SW (2000) Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* 29:213-238.
60. Roder H, Elove GA, & Englander SW (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335(6192):700-704.
61. Udgaonkar JB & Baldwin RL (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335(6192):694-699.
62. Bycroft M, Matouschek A, Kellis JT, Jr., Serrano L, & Fersht AR (1990) Detection and characterization of a folding intermediate in barnase by NMR. *Nature* 346(6283):488-490.
63. Lu J & Dahlquist FW (1992) Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry* 31(20):4749-4756.
64. Matouschek A, Serrano L, Meiering EM, Bycroft M, & Fersht AR (1992) The folding of an enzyme. V. H/2H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: complementarity to and agreement with protein engineering studies. *J Mol Biol* 224(3):837-845.
65. Radford SE, Dobson CM, & Evans PA (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358(6384):302-307.
66. Jennings PA & Wright PE (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262(5135):892-896.
67. Mullins LS, Pace CN, & Raushel FM (1993) Investigation of ribonuclease T1 folding intermediates by hydrogen-deuterium amide exchange-two-dimensional NMR spectroscopy. *Biochemistry* 32(24):6152-6156.
68. Jacobs MD & Fox RO (1994) Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. *Proc Natl Acad Sci U S A* 91(2):449-453.
69. Raschke TM & 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(4):298-304.
70. Bollen YJ, Kamphuis MB, & van Mierlo CP (2006) The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* 103(11):4095-4100.
71. Krishna MM, Lin Y, Rumbley JN, & Englander SW (2003) Cooperative omega loops in cytochrome c: role in folding and function. *J Mol Biol* 331(1):29-36.
72. Vadrevu R, Wu Y, & Matthews CR (2008) NMR analysis of partially folded states and persistent structure in the alpha subunit of tryptophan synthase: implications for the equilibrium folding mechanism of a 29-kDa TIM barrel protein. *J Mol Biol* 377(1):294-306.
73. Yan S, Kennedy SD, & Koide S (2002) Thermodynamic and kinetic exploration of the energy landscape of Borrelia burgdorferi OspA by native-state hydrogen exchange. *J Mol Biol* 323(2):363-375.
74. Miranker A, Radford SE, Karplus M, & Dobson CM (1991) Demonstration by NMR of folding domains in lysozyme. *Nature* 349(6310):633-636.
75. Kato H, Vu ND, Feng H, Zhou Z, & Bai Y (2007) The folding pathway of T4 lysozyme: an on-pathway hidden folding intermediate. *J Mol Biol* 365(3):881-891.

76. Bai Y, Milne JS, Mayne L, & Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17(1):75-86.
77. Schmid FX & Baldwin RL (1979) Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange. *J Mol Biol* 135(1):199-215.
78. Yamasaki K, Ogasahara K, Yutani K, Oobatake M, & Kanaya S (1995) Folding pathway of Escherichia coli ribonuclease HI: a circular dichroism, fluorescence, and NMR study. *Biochemistry* 34(51):16552-16562.
79. Roder H & Wuthrich K (1986) Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins* 1(1):34-42.
80. Gladwin ST & Evans PA (1996) Structure of very early protein folding intermediates: new insights through a variant of hydrogen exchange labelling. *Fold Des* 1(6):407-417.
81. Kim PS & Baldwin RL (1980) Structural intermediates trapped during the folding of ribonuclease A by amide proton exchange. *Biochemistry* 19(26):6124-6129.
82. Berger A & Linderstrom-Lang K (1957) Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* 69:106-118.
83. Llinas M, Gillespie B, Dahlquist FW, & Marqusee S (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* 6(11):1072-1078.
84. Chu R, Pei W, Takei J, & Bai Y (2002) Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. *Biochemistry* 41(25):7998-8003.
85. Arrington CB & Robertson AD (1997) Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry* 36(29):8686-8691.
86. Sivaraman T, Arrington CB, & Robertson AD (2001) Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. *Nat Struct Biol* 8(4):331-333.
87. Yi Q & Baker D (1996) Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. *Protein Sci* 5(6):1060-1066.
88. Rodriguez HM, Robertson AD, & Gregoret LM (2002) Native state EX2 and EX1 hydrogen exchange of Escherichia coli CspA, a small beta-sheet protein. *Biochemistry* 41(7):2140-2148.
89. Hoang L, Bedard S, Krishna MM, Lin Y, & Englander SW (2002) Cytochrome c folding pathway: kinetic native-state hydrogen exchange. *Proc Natl Acad Sci U S A* 99(19):12173-12178.
90. Cliff MJ, Higgins LD, Sessions RB, Waltho JP, & Clarke AR (2004) Beyond the EX1 limit: probing the structure of high-energy states in protein unfolding. *J Mol Biol* 336(2):497-508.
91. Ferraro DM, Lazo ND, & Robertson AD (2004) EX1 hydrogen exchange and protein folding (vol 43, pg 587, 2004). *Biochemistry* 43(12):3756-3756.
92. Schanda P, Forge V, & Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc Natl Acad Sci U S A* 104(27):11257-11262.
93. Wani AH & Udgaonkar JB (2009) Native state dynamics drive the unfolding of the SH3 domain of PI3 kinase at high denaturant concentration. *Proc Natl Acad Sci U S A* 106(49):20711-20716.

94. Silverman JA & Harbury PB (2002) Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J Biol Chem* 277(34):30968-30975.
95. Cliff MJ, *et al*. (2006) A thiol labelling competition experiment as a probe for sidechain packing in the kinetic folding intermediate of N-PGK. *J Mol Biol* 364(4):810-823.
96. Feng Z, Butler MC, Alam SL, & Loh SN (2001) On the nature of conformational openings: native and unfolded-state hydrogen and thiol-disulfide exchange studies of ferric aquomyoglobin. *J Mol Biol* 314(1):153-166.
97. Ha JH & Loh SN (1998) Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat Struct Biol* 5(8):730-737.
98. Isom DG, Vardy E, Oas TG, & Hellinga HW (2010) Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity. *Proc Natl Acad Sci U S A* 107(11):4908-4913.
99. Jha SK & Udgaonkar JB (2007) Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* 282(52):37479-37491.
100. Johnson CP, Tang HY, Carag C, Speicher DW, & Discher DE (2007) Forced unfolding of proteins within cells. *Science* 317(5838):663-666.
101. Sridevi K & Udgaonkar JB (2002) Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry* 41(5):1568-1578.
102. Krishnan B & Gierasch LM (2011) Dynamic local unfolding in the serpin alpha-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nat Struct Mol Biol* 18(2):222-226.
103. Silverman JA & Harbury PB (2002) The equilibrium unfolding pathway of a (beta/alpha)8 barrel. *J Mol Biol* 324(5):1031-1040.
104. Dabora JM & Marqusee S (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: characterization of a partially folded state. *Protein Sci* 3(9):1401-1408.
105. Connell KB, Miller EJ, & 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(2):450-460.
106. Cecconi C, Shank EA, Bustamante C, & Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. *Science* 309(5743):2057-2060.
107. Ishikawa K, *et al*. (1993) Crystal structure of ribonuclease H from Thermus thermophilus HB8 refined at 2.8 A resolution. *J Mol Biol* 230(2):529-542.
108. Hollien J & Marqusee S (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* 38(12):3831-3836.
109. Hollien J & Marqusee S (2002) Comparison of the folding processes of T. thermophilus and E. coli ribonucleases H. *J Mol Biol* 316(2):327-340.
110. Hollien J & Marqusee S (1999) Structural distribution of stability in a thermophilic enzyme. *Proc Natl Acad Sci U S A* 96(24):13674-13678.
111. Ratcliff K & Marqusee S (2010) Identification of residual structure in the unfolded state of ribonuclease H1 from the moderately thermophilic Chlorobium tepidum: comparison with thermophilic and mesophilic homologues. *Biochemistry* 49(25):5167-5175.

112. Robic S, Guzman-Casado M, Sanchez-Ruiz JM, & Marqusee S (2003) Role of residual structure in the unfolded state of a thermophilic protein. *Proc Natl Acad Sci U S A* 100(20):11345-11349.
113. Guzman-Casado M, Parody-Morreale A, Robic S, Marqusee S, & Sanchez-Ruiz JM (2003) Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of Thermus thermophilus ribonuclease H. *J Mol Biol* 329(4):731-743.
114. Young TA, Skordalakes E, & Marqusee S (2007) Comparison of proteolytic susceptibility in phosphoglycerate kinases from yeast and E. coli: modulation of conformational ensembles without altering structure or stability. *J Mol Biol* 368(5):1438-1447.

Chapter 2

Exploring Subdomain Cooperativity in T4 Lysozyme: Uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway

Adapted from Cellitti J, Bernstein R, & Marqusee S (2007) *Prot Sci* 16(5):852-862

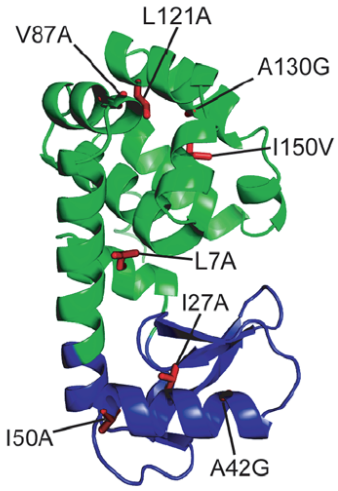
2.1 Introduction

Partially folded forms of proteins are instrumental in a wide variety of important biological functions such as folding (1-3), allostery (4, 5), catalysis (6, 7), protein translocation (8), and signal transduction (9, 10), as well as unwanted processes such as aggregation and formation of amyloid fibrils (11, 12). Understanding the structure, energetics, and dynamics of such intermediate states is thus a key problem in modern biochemistry. The rare and ephemeral nature of such species, however, makes them extremely difficult to study – with a lifetime often less than milliseconds, at equilibrium they typically only represent about 1 in 10^6 molecules.

Experimental techniques using amide hydrogen exchange (HX) have proven extremely useful for investigating such elusive species. Transiently populated folding intermediates can be labeled by pulse-labeling HX studies, while native state HX (NSHX) experiments yield information about rarely populated forms at equilibrium with the native state during the experimental time period. For most proteins investigated by both the kinetic and equilibrium approaches, the two yield similar results, suggesting that they report on the same intermediate. The bacteriophage protein T4 lysozyme is one notable exception.

T4 lysozyme, or T4L*(* refers to a cysteine-free variant (13)), is a small (164 residue) protein that displays cooperative unfolding behavior with an equilibrium denaturation profile well-modeled by a two-state transition. This simple picture, however, is complicated by structural and HX data. Although considered to be a single domain, T4 lysozyme is actually composed of two structural lobes referred to as the N-terminal and C-terminal subdomains (see Figure 1). NSHX has revealed a high energy partially unfolded form that appears structured in only the C-terminal subdomain (14), while pulse-labeling HX studies show protection factors that implicate an early intermediate with structure in both the N- and C-terminal subdomains. Hence, for T4 lysozyme, the transient folding intermediate detected by pulse-labeling HX does not mirror the high energy intermediate detected by NSHX.

Figure 1. T4 lysozyme and the sites of mutations studied; the C-terminal subdomain is shown in green and the N-terminal subdomain is shown in blue.



A prior investigation focused on structural and energetic characterizations of the partially unfolded form detected by NSHX (15). These studies demonstrated that a peptide model of this

intermediate that contains residues from just the C-terminal subdomain forms a cooperatively folded native-like structure. Thus, residues in the N-terminal subdomain are not required to direct the folding of the C-terminal subdomain. These studies also demonstrated that in the full-length protein there is cooperative coupling between the two lobes brought about, in part, by helix A. Although this helix resides sequentially at the N-terminus of the protein, structurally it contributes to the C-terminal subdomain. Circular permutation (placing the N-terminal helix at the C-terminus of the protein) diminishes the cooperativity between the subdomains, resulting in an even greater population of molecules with an independently folded C-terminal lobe.

Previous reports on the folding kinetics of T4L* have noted that the chevron plot (ln k_{obs} versus denaturant) displays the characteristic rollover in low denaturant suggestive of an early folding intermediate (16). This kind of rollover can be modeled by either an on-pathway or an off-pathway early intermediate (U ⇆ I ⇆ N or I ⇆ U ⇆ N). For simplicity, the data for most proteins (including T4L*) are fit to the on-pathway obligatory folding intermediate model. Other refolding studies have also suggested the presence of a potential folding intermediate based on mutational analyses and detailed thermodynamic characterization of the refolding reaction ((17-21).

Here, we use a protein engineering approach in conjunction with stopped-flow circular dichroism (CD) and chevron analysis to further characterize the transient intermediates populated during the folding and unfolding processes of the protein. We find that the chevron plot of T4L* displays rollover not only in the folding limb but also in the unfolding limb, a phenomenon that has been observed (16, 19) but never interpreted for this protein. By analyzing mutations in both structural subdomains of the protein (Figure 1), in addition to CP13* (22) and the LCTF variants introduced in (15), we suggest that the intermediate probed by NSHX resides on the native state side of the rate-limiting transition state for folding and thereby represents an unfolding intermediate. Such intermediates, while difficult to identify, represent conformations easily sampled during excursions from the native state and can play an important role in both folding and misfolding processes. Our conclusions also resolve a long-standing discrepancy between kinetic and equilibrium hydrogen exchange results for T4L* and highlight the importance of using multiple tools and approaches to characterize the energy landscape of a protein.

2.2 Methods

2.2.1 Site-directed mutagenesis

Mutants were created using the QuikChange method. Mutations were chosen and primers designed by Eric Nicholson.

2.2.2 Protein purification

All proteins were overexpressed and purified as previously described (22).

2.2.3 CD spectrometry and denaturation melts

Spectra and denaturation melts on all mutants were done as described for T4L* (22).

2.2.4 Kinetic folding and unfolding

Refolding was initiated by diluting unfolded protein at 4.0 mg/ml in urea (6.5 M for T4L*, 6.0 M for mutants) 1:11 (v/v) into refolding buffer (20 mM potassium phosphate, pH6.0; 100 mM NaCl

with low concentrations of urea) using an AVIV 202 stopped-flow CD spectrophotometer equipped with a delta mixer and a cuvette pathlength of 1 mm. The CD signal was followed over time at 222 nm. Unfolding experiments were performed by diluting folded protein in 1.0 M urea into high concentrations of urea. Folded and unfolded stock protein samples were equilibrated overnight. Each data point represents the averaged signal of at least seven individual shots. Experiments were performed at 25 °C.

The observed signal for each concentration of urea described a single exponential (k_{obs}) and was fit using SigmaPlot 8.0 for Windows (Systat Software Inc., Richmond, CA) to the equation:

$$\text{Signal} = A \exp(-k_{obs}t) + C$$

where C is the final signal, A the amplitude of the observable phase, and t is time. The initial points in the "rolled over" limbs of each variant describe the "burst phase" (the signal obtained in the undetected dead-time of the instrument, 18.1 ms), measured by:

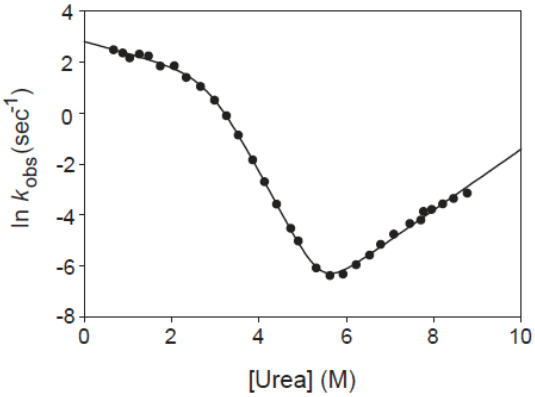
$$A_{bp} = (C + A) - \text{unfolded signal}$$

2.3 Results

*2.3.1 Folding kinetics for T4L**

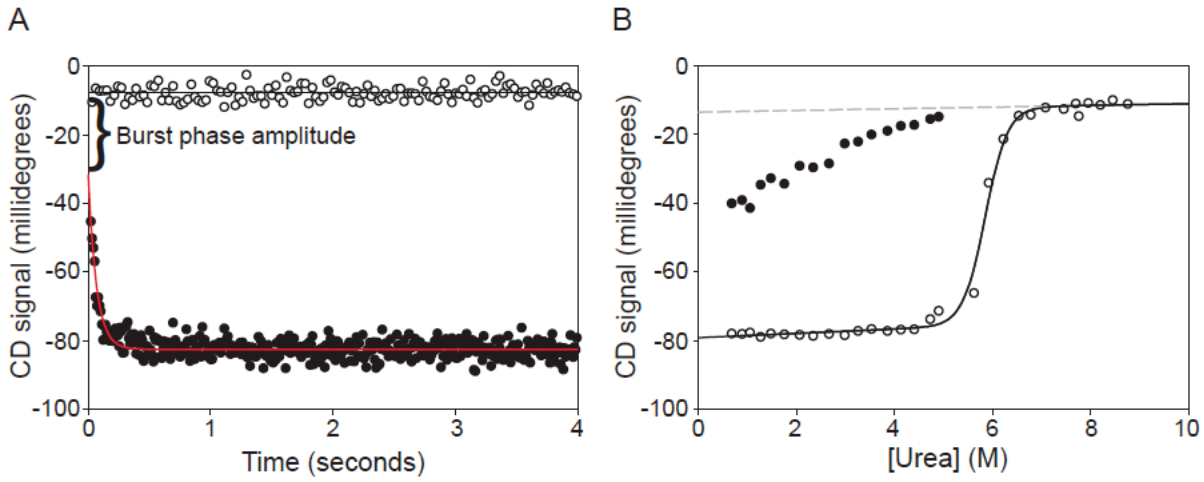
The unfolding/refolding process of T4L* was monitored as a function of urea concentration by following the change in CD signal as a function of time after rapid dilution from urea using a stop-flow mixing device. At all concentrations of urea, the observed data were fit to a single exponential process. A chevron plot derived from these data is shown in Figure 2. As seen previously (16, 19), the chevron plot for T4L* shows a rollover at low denaturant concentrations.

Figure 2. Chevron plot of T4L*. The solid line is a fit using the three-state, on-pathway model.



At low denaturant concentrations, in the rollover region of the chevron plot, the amplitude of the observed phase does not account for the overall signal change of the reaction. Figure 3 shows how both these so-called burst phase amplitudes and the final signals vary as a function of denaturant. Unlike the final signal, which reveals a cooperative unfolding transition, the burst-phase amplitudes change in a more-or-less linear fashion.

Figure 3. Burst phase amplitudes for T4L*. (A) Determining the burst phase amplitude at a final urea concentration of 0.6M. The difference in signal between the unfolded protein (○) the initial signal upon refolding (●) extrapolated to 0 time is the burst phase amplitude. The red line is a fit of the data to a single exponential while the solid black line represents the average unfolded signal. (B) The burst phase as a function of urea. Burst phase amplitudes (●) compared to final amplitudes (○). The solid black line represents a fit of final amplitudes to a two-state sigmoidal transition while the gray dashed line is an extrapolation of the unfolded baseline to low denaturant.



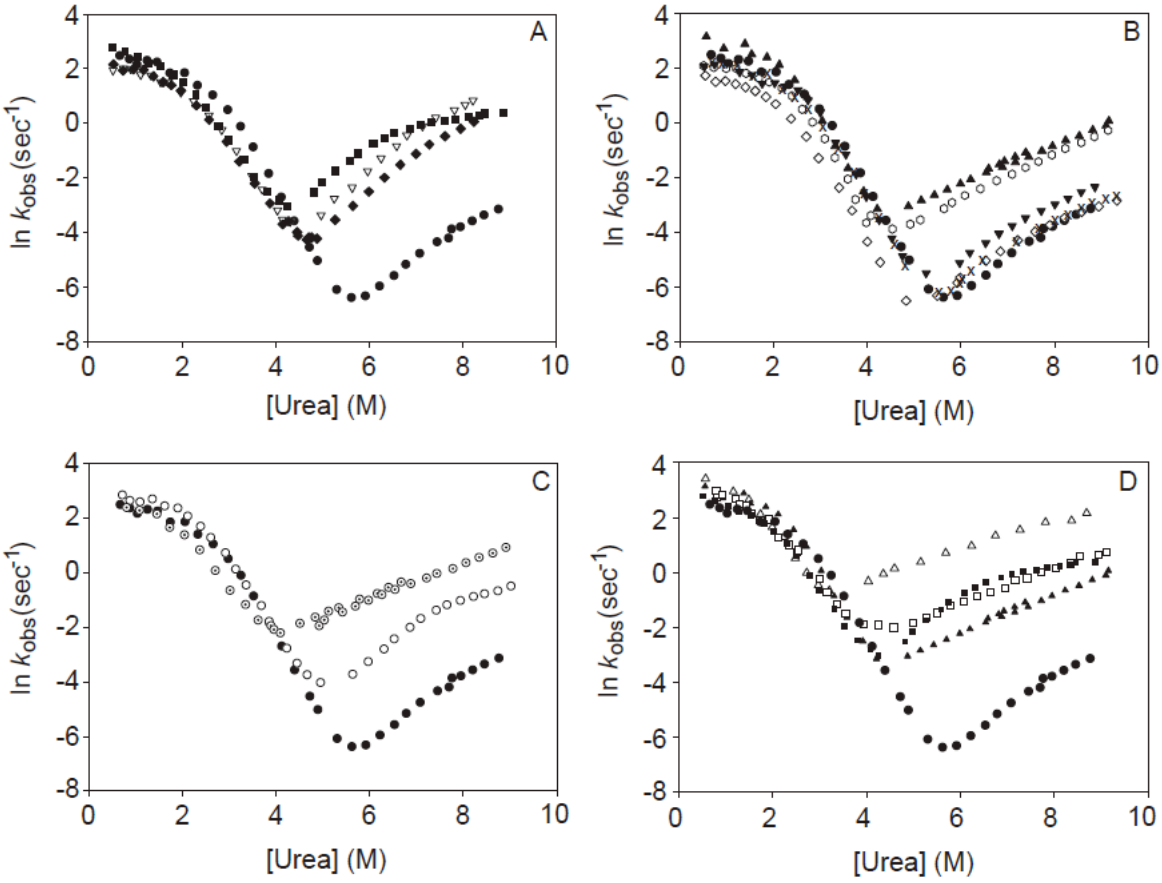
2.3.2 Folding of T4L variants*

Several site-specific variants were generated in both subdomains of T4L*: I27A, A42G, and I50A in the N-terminal subdomain and L7A, V87A, L121A, A130G, and I150V in the C-terminal subdomain. Mutations were designed to be minimally disruptive of structure and were placed at positions used as probes in HX studies. Most of the variants result in the removal of small hydrophobic side chains. All the mutant proteins display apparent two-state behavior at equilibrium, and the resulting stabilities derived from the two-state fit are reported in Table 1. The refolding/unfolding kinetics were monitored by CD in a manner similar to that described for T4L*, and the resulting chevron plots are shown in Figure 4.

Table 1. Stabilities of T4L*, its mutants and variants. $\Delta\Delta G°$ is the difference in stability of the mutants and T4L* except numbers in parenthesis, which represent $\Delta\Delta G°$ values between CP13* and the mutants.

	ΔG° (kcal/mol)	*m*-value (kcal/mol M^{-1})	ΔΔG° (kcal/mol)
T4L*	14.9 ± 0.5	2.7 ± 0.1	
N-terminal variants			
I27A	11.0 ± 0.2	2.6 ± 0.1	3.9
A42G	11.7 ± 0.3	2.6 ± 0.1	3.2
I50A	12.0 ± 0.3	2.6 ± 0.1	2.9
C-terminal variants			
L7A	10.8 ± 0.4	2.7 ± 0.1	4.1
V87A	12.1 ± 0.4	2.5 ± 0.1	2.8
L121A	10.7 ± 0.2	2.6 ± 0.1	4.2
A130G	12.8 ± 0.3	2.6 ± 0.1	2.1
I150V	12.8 ± 0.4	2.5 ± 0.1	2.1
Large-scale variants			
LCTF	6.8 ± 0.4	1.8 ± 0.1	8.1
CP13*	11.2 ± 0.3	2.3 ± 0.1	3.7
CP13* L7A	7.5 ± 0.3	2.3 ± 0.1	7.4 (3.7)
CP13* I27A	8.8 ± 0.4	2.2 ± 0.1	6.1 (2.4)

Figure 4. Chevrons of T4L* mutants. (A) N-terminal mutations. T4L* (●), I27A (▪), A42G (▽), and I50A (◆). (B) C-terminal mutations. T4L* (●), L7A (▲), V87A (◇), L121A (⬡), A130G (▼), and I150V (**X**). (C) Large-scale variants. T4L* (●), CP13* (○), and LCTF (⊙). (D) Mutants in the CP13* background. I27A (▫) and L7A (Δ) in the context of the circular permutant. T4L* (●) and the mutations in the wild type background, L7A (▲) and I27A (▪), are shown for clarity.



Chevron plots for the three variants containing point mutations in the N-terminal subdomain – I27A, A42G, and I50V – are shown in Figure 4A. These mutations all increase the rate of unfolding with minimal effects on the refolding rate. Thus, faster unfolding accounts for the decreased stability of these mutations. The other notable observation is the dramatic downward curvature in the unfolding limb of I27A, present to a lesser extent in the other two variants as well. These data were best fit to a four state model, described in detail below.

Chevron plots for the five variants in the C-terminal subdomain are shown in Figure 4B. As with the N-terminal mutations, the folding limb is minimally affected. Three of the C-terminal mutations show only small effects on the unfolding limb. On the other hand, the unfolding rates of the two remaining variants, L7A and L121A, are markedly increased and perhaps most interestingly, their unfolding limbs appear to be entirely devoid of unfolding-limb curvature and were fit adequately using a three state model (see below).

Refolding kinetics of the circular permutant CP13* were also monitored by stop-flow circular dichroism. CP13* has residues 1-12 grafted onto its C-terminus using a six amino acid Ser$(Gly)_4$Ala linker such that helix A is both sequentially and structurally associated with the rest of the C-terminal subdomain. CP13* retains a wild-type-like structure with minimal change to the overall stability (Table 1) (15). Compared to T4L*, however, the energetics of the two subdomains are less coupled in CP13*, as shown by NSHX (15). The chevron plot for this variant is shown in Figure 4C (circles). Like the other variants presented above, the refolding limb is minimally affected by this permutation and the destabilization appears to be reflected almost exclusively in the faster unfolding of the protein. It should also be noted that the unfolding limb of this permutation displays marked curvature.

LCTF (long C-terminal fragment), as described in (15), was designed as a fragment modeling the partially unfolded form detected at equilibrium using NSHX. This fragment contains the C-terminal region of the protein, starting at residue 60, followed by the Ser$(Gly)_4$Ala linker described for CP13*, and residues 1-12 that constitute helix A. Essentially, LCTF represents the C-terminal 98 residues of CP13* (the C-terminal structural subdomain). LCTF forms a cooperatively folded domain (Table 1) with a structure that mirrors the C-terminal subdomain of the full-length protein(15). The chevron plot for the folding kinetics of LCTF is shown in Figure 4C; the plot looks remarkably like those for the full-length variants. The refolding rate appears very similar to that of the wild type, CP13*, and the other variants, with the rollover at low denaturant concentrations. Unlike CP13*, however, LCTF shows no apparent curvature in its unfolding limb and was fit adequately to a three state model (see below).

Finally, we evaluated the folding of the L7A and I27A mutations in the context of CP13* (Figure 4D). Similar to mutations in the wild type background, these mutations in CP13* have little effect on the folding limb and its rollover. On the other hand, the effects of these mutations on the unfolding limbs in the CP13* background are opposite from their effects in the wild type background. While L7A eliminates unfolding limb curvature in the T4L* protein, the curvature remains in the CP13* background. Similarly, the greater unfolding limb curvature induced by I27A in T4L* is eliminated entirely when the mutation is in the CP13* context.

2.3.3 Analysis of the chevron plots

All of the kinetic data from this study were initially fit to a three-state model using equations 1-5 (23).

$$\ln k_{obs} = \ln (f_I \cdot k_{IN} + k_{NI}) \quad (1)$$

$$f_I = K_{UI} / (1 + K_{UI}) \quad (2)$$

$$k_{IN}^{(DEN)} = k_{IN}^{(H2O)} \cdot \exp (-m_{IN} \cdot x / RT) \quad (3)$$

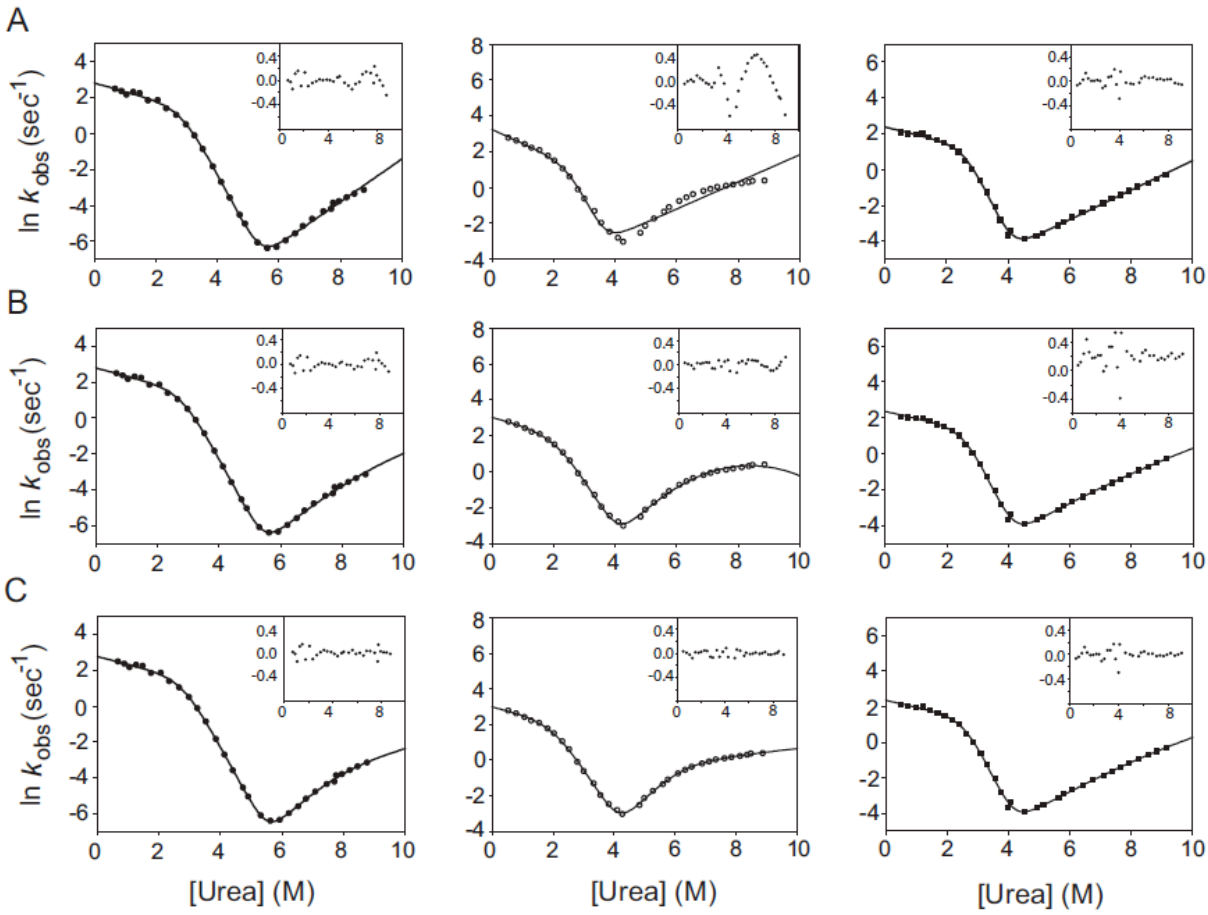
$$k_{NI}^{(DEN)} = k_{NI}^{(H2O)} \cdot \exp (-m_{NI} \cdot x / RT) \quad (4)$$

$$K_{UI}^{(DEN)} = K_{UI}^{(H2O)} \cdot \exp (-m_{UI} \cdot x / RT) \quad (5)$$

where $k_{IN}^{(DEN)}$ and $k_{NI}^{(DEN)}$ are the microscopic rate constants for folding and unfolding, respectively, $k_{IN}^{(H2O)}$ and $k_{NI}^{(H2O)}$ the rate constants in water, K_{UI} the equilibrium constant between U and I, and f_I the fraction of molecules present as I. m_{IN}, m_{NI} and m_{UI} are the denaturant dependencies of the two rate constants and equilibrium constant, respectively. x is denaturant concentration, R the gas constant in kcal/mol·K, and T the temperature in Kelvin. Residuals

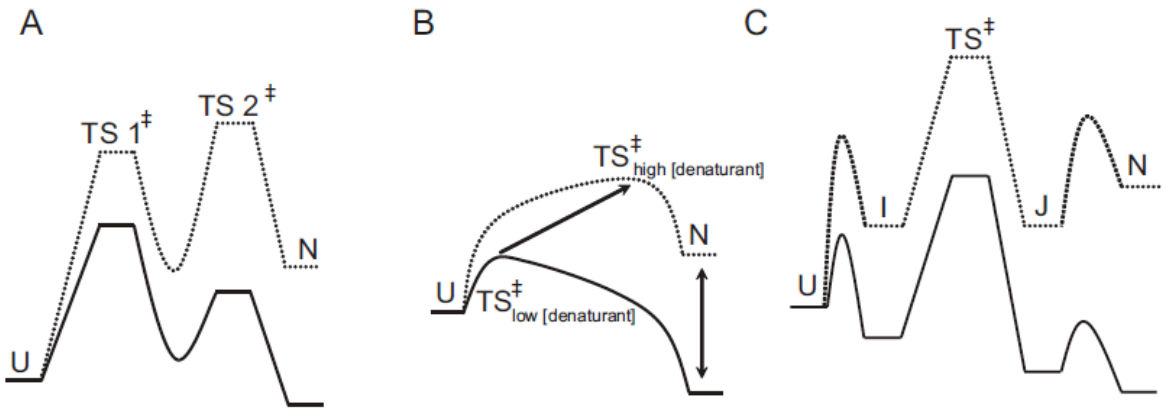
from these fits, particularly from unfolding limb data, demonstrate the poorness of this model for all but three of the proteins. These fits and the residuals based on this model are shown in Figure 5A for T4L*, I27A and L121A.

Figure 5. Fits of T4L* and representative mutants. T4L* (●), I27A (○), and L121A (■). Residuals of all fits are inset. (A) Three-state fit; (B) Hammond behavior; (C) Four-state model.



Faced with rate profiles more complex than a simple three-state, on-pathway intermediate, alternative explanations for the curved chevron limbs were investigated. Deviations from linearity in chevron plots are typically attributed to one of two denaturant-dependent effects: changes in the barriers (transition state) or changes in local minima, including any low energy populated states on either side of the rate limiting step. The two most common interpretations of curvature in the unfolding limb deal with effects on the folding barriers: 1) the moving transition state model, or Hammond behavior, (24), and 2) the high energy intermediate model (25-28). An alternate explanation is a four-state model that invokes both folding and unfolding intermediates. Reaction coordinate diagrams illustrating these models at high and low denaturant concentrations are shown in Figure 6.

Figure 6. Theoretical reaction coordinates representing the most common interpretations of complex unfolding and refolding kinetics. Solid lines represent low [denaturant] conditions, dotted lines are at high [denaturant]. (A) Transition state switching; the stability of the high energy intermediate cannot be measured. (B) Hammond behavior; transition state "sliding". (C) Four-state model; the barriers between U and I and N and J cannot be measured.



The moving transition state model is based on the idea that the transition state in protein folding is a broad continuum of conformers referred to as the transition state ensemble (TSE) (29-31). In this broad continuum, one set of states will be rate limiting under a given set of conditions, but by changing solvent, i.e., increasing denaturant, or by altering the protein through mutation, the highest energy state in this continuum will "slide" along the TSE (Figure 6A). In the absence of denaturant, the barrier will be closer on the reaction coordinate to the unfolded state, but with increasing denaturant, the transition state will become more native-like. Including a denaturant dependence term for this barrier is used to account for such Hammond behavior. This model has been successfully applied to chymotrypsin inhibitor 2 (29), barnase at different temperatures (32), and most interestingly in the case of two structural homologs U1A and S6 (31). The revised rate equations are:

$$k_{IN}^{(DEN)} = k_{IN}^{(H2O)} \cdot \exp\left(-a_{IN} \cdot [Den]/RT + b_{IN} \cdot [Den]^2/RT\right) \quad (6)$$

$$k_{NI}^{(DEN)} = k_{NI}^{(H2O)} \cdot \exp\left(-a_{NI} \cdot [Den]/RT + b_{NI} \cdot [Den]^2/RT\right) \quad (7)$$

and, as the derivative of ln k with respect to denaturant, the actual m-value becomes:

$$m_{XY} = a_{XY} + 2b_{XY} \cdot [Den] \quad (8)$$

This approach, however, did not adequately fit the data for T4L*. Fits and residuals for this model to T4L*, I27A, and L121A are shown in Figure 5B. Fitting the data to this model, modified slightly to include an on-pathway folding intermediate, resulted in rate constants that correspond to a dramatic overestimation of ΔG_{unf}. The existence of T4L* mutants that eliminate the curvature in the unfolding limb is also difficult to reconcile with this model.

The second model invoking effects on the transition state was developed by Kiefhaber and co-workers (26, 27, 33-35) and has been applied to a number of proteins (25, 28, 36, 37). This model can be thought of as an extreme case of a single broad transition state in which there are

two local high energy maxima. The two transition states can be differentially affected by denaturant so that the first transition state (TS1) is rate limiting at low denaturant while the second transition state (TS2) is rate limiting at high denaturant concentrations.

We have so far been unable to fit our data with this model. However, given the dramatic differences in the unfolding limb profiles among our variants, it is unlikely that the changes are due simply to effects on the transition state. In this model, a mutation that destabilizes the native state would also have to stabilize TS2 to prevent it from becoming rate-limiting at high denaturant. Even if this were possible, the effect upon TS2 would have to be so dramatic as to almost eliminate it. If not, the unfolding kinetics we observe in variants without curvature in the unfolding limb would be biphasic since there would be two significant barriers.

Finally, the data were fit using a four-state model with on-pathway folding and unfolding intermediates (38) (Scheme 1).

$$U \overset{K_{UI}}{\rightleftharpoons} I \underset{k_{UI}}{\overset{k_{IJ}}{\rightleftharpoons}} J \overset{K_{NJ}}{\rightleftharpoons} N \qquad \text{(Scheme 1)}$$

This model is identical to the three-state model with the exception that a term for the fraction of protein present as an unfolding intermediate, J, must be included.

$$\ln k_{obs} = \ln (f_I \cdot k_{IN} + f_J \cdot k_{NJ}) \qquad (9)$$

$$f_J = K_{UJ} / (1 + K_{UJ}) \qquad (10)$$

$$K_{UJ}^{(DEN)} = K_{UJ}^{(H2O)} \cdot \exp (-m_{UJ} \cdot x / RT) \qquad (11)$$

Figure 5C shows fits of the data from T4L*, I27A, and L121A to this model with the residuals inset. Variants such as L121A devoid of curvature in the unfolding limb can be fit with this model but the resulting K_{NJ} values indicate that J never becomes populated. Thus, the simpler three-state on-pathway model was used for these variants. F-tests on the residuals of the three- and four- state fits confirm that for all but L121A, L7A, and LCTF, the four-state fit is significantly better (data not shown). Results for the kinetic fits of the chevrons are shown in Table 2.

Table 2. Kinetic parameters from a four state fit of T4L*, its mutants and variants, except where noted above.

	k_{ij} (sec^{-1})	m_{ij} (kcal/mol M^{-1})	K_{UI}	m_{UI} (kcal/mol M^{-1})	k_{JI} (sec^{-1})	m_{JI} (kcal/mol M^{-1})	K_{NJ}	m_{NJ} (kcal/mol M^{-1})
T4L*	15.8 (± 1.3)	0.3 (± 0.04)	1926 (± 437)	1.6 (± 0.03)	5.0E-4 (± 1.9E-3)	-0.3 (± 0.2)	1.3E-4 (± 3.0E-4)	-0.7 (± 0.09)
N-terminal variants								
I27A	20.0 (± 1.4)	0.3 (± 0.04)	113 (± 24)	1.4 (± 0.03)	0.25 (± 0.08)	-0.1 (± 0.02)	4.0E-5 (± 1.1E-5)	-1.0 (± 0.04)
A42G	9.5 (± 0.8)	0.21 (± 0.05)	296 (± 90)	1.5 (± 0.04)	5.1E-2 (± 0.1)	-0.3 (± 0.1)	7.3E-05 (±1.2E-4)	-0.8 (±0.1)
I50A	9.5 (± 1.0)	0.1 (± 0.07)	103 (± 23)	1.4 (± 0.05)	4.7E-3 (± 8.5E-3)	-0.4 (±0.1)	1.6E-4 (± 2.0E-4)	-0.8 (± 0.06)
C-terminal variants								
L7A	31.8 (± 4.3)	0.4 (± 0.1)	2457 (± 766)	1.8 (± 0.1)	1.8E-3 (± 1.2E-4)	-0.4 (± 0.01)		
V87A	6.8 (± 0.4)	0.2 (0.03)	575 (± 92)	1.6 (± 0.03)	5.17E-4 (± 4.3E-4)	-0.3 (± 0.1)	2.7E-04 (±1.2E-4)	-0.7 (± 0.1)
L121A	10.7 (± 1.0)	0.3 (± 0.04)	5251 (± 2326)	1.9 (± 0.1)	4.3E-4 (± 1.0E-4)	0.2 (± 0.04)		
A130G	10.4 (± 1.0)	0.2 (± 0.04)	4393 (± 1683)	1.8 (± 0.05)	5.4E-3 (± 0.04)	-0.2 (± 0.4)	6.5E-4 (± 3.7E-3)	-0.5 (± 0.2)
I150V	11.4 (± 1.4)	0.3 (± 0.04)	1056 (± 281)	1.5 (± 0.04)	1.5E-4 (± 1.5E-4)	-0.4 (± 0.1)	2.5E-5 (± 4.7E-5)	-1.0 (± 0.2)
Large-scale variants								
LCTF	16.4 (± 4.2)	0.3 (± 0.1)	200 (± 131)	1.5 (± 0.1)	7.1E-3 (± 0.8E-3)	-0.4 (± 0.01)		
CP13*	17.7 (± 1.6)	0.1 (± 0.04)	401 (± 65)	1.5 (± 0.03)	0.19 (± 0.25)	-0.1 (± 0.09)	1.1E-5 (± 9.1E-6)	-0.9 (± 0.04)
CP13* L7A	48.1 (± 7.8)	0.4 (± 0.1)	339 (± 379)	1.7 (± 0.2)	0.9 (± 2.5)	-0.2 (± 0.2)	3.8E-2 (± 7.8E-2)	-0.3 (± 0.1)
CP13* I27A	31.9 (± 5.3)	0.4 (± 0.1)	97 (± 53)	1.3 (± 0.1)	7.1E-3 (± 8.0E-4)	-0.4 (± 0.01)		

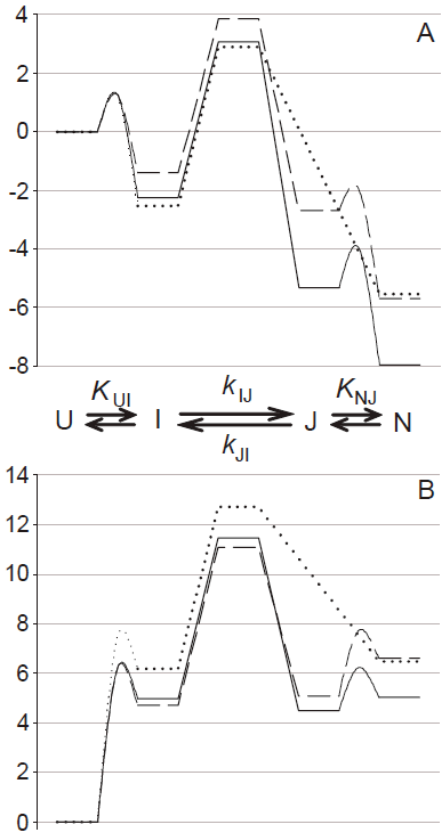
Based on the results from the above kinetic models, the stabilities of the intermediates and the activation energies of the major rate-limiting transition state were calculated using the following:

$$\Delta G^{\ddagger} = -RT \ln (k_{XY}/k_0) + m \cdot x / RT \qquad (12)$$

$$\Delta G_{intermediate} = -RT \ln K_{XY} + m \cdot x / RT \qquad (13)$$

Although the absolute height of the energy barriers is not important, a value of k_0 was set at 10^9 sec^{-1} (39). Reaction-coordinate diagrams for T4L*, I27A, and L121A are shown at both 0 and 9M urea in Figure 7.

Figure 7. Reaction coordinate diagrams for T4L* (━), I27A (---), and L121A (···). Diagrams are normalized to U which is assigned 0. Barriers between U and I and N and J cannot be determined were arbitrarily chosen. (A) 0M. The native states of both mutants are destabilized but only I27A is significantly destabilized in I and the transition state. (B) 9M. J is more destabilized than N in I27A, making it a more populated state in this mutant than in T4L*.



2.4 Discussion

T4 lysozyme has been reported to fold through an early intermediate based on several observations: significant burst phase amplitudes, rollover in the folding limb of the chevron plot, and pulse-labeling HX experiments that demonstrate a small amount of protection for a subset of amides. Based on the distribution of protected residues, both the N-and C-terminal subdomains are involved in this intermediate (40). In carrying out a detailed chevron analysis, we now find evidence for an intermediate on the folded side of the rate-limiting transition state which for many of the variants studied here results in the population of an unfolding intermediate at high urea concentrations.

Many of our chevron plots revealed substantial curvature in the unfolding limb which cannot be accounted for by a simple three-state, on-pathway intermediate model. By invoking changes in the ground states we have found a model with two intermediates, one populated upon folding (I) and one upon unfolding (J), that adequately explains our data. Single point mutations can either accentuate the population of the intermediate or eliminate it. A mutation that destabilizes N may also destabilize J; however, a destabilizing mutation that resides in a region not structured in J

may actually stabilize J relative to N. Hence an increase in the relative curvature in the unfolding limb would be expected to correspond to changes in regions not involved in the structure and stability of the unfolding intermediate.

A comparison of the effects on folding between the different variants suggests that this unfolding intermediate corresponds to the partially unfolded form detected by NSHX that is structured only in the C-terminal domain. Mutations that specifically destabilize the N-terminal subdomain, such as I27A, stabilize this intermediate, resulting in greater rollover of the unfolding arm of the chevron plot, while mutations that destabilize the C-terminal subdomain either have no impact on the intermediate's formation because of equal effects on the native and high energy states, as seen in V87A, A130G, and I150V, or they destabilize the intermediate to such an extent that it does not form at all, leading to chevron plots with linear unfolding arms for L7A and L121A. In both cases, it is the mutations with the largest effect on the free energy of unfolding that have the greatest impact on the intermediate's formation or disruption.

The chevron plots of the two large scale variants of T4L*, CP13* and LCTF, also support the idea that the unfolding intermediate resembles that detected by NSHX and modeled by LCTF. CP13*, shown by native state hydrogen exchange to specifically destabilize the N-terminal subdomain(15), has a notably increased curvature in the unfolding limb of its chevron plot compared to T4L*, suggesting that the unfolding intermediate does not result from structure or stability in the N-terminal domain. In LCTF, the entire N-terminal subdomain is absent, as is any curvature of the chevron's unfolding limb. This too is consistent with the notion that the C-terminal region models the unfolding intermediate. Furthermore, the extrapolated folding rate of LCTF is similar to that of the intact protein, suggesting that for both variants folding may involve crossing a similar rate-limiting transition state. Hence, the kinetic profile of LCTF further supports the hypothesis that the intermediate containing an unfolded N-terminal domain and a folded C-terminal domain, as detected by NSHX, resides on the folded side of the rate limiting transition state.

Mutations made in the context of the circular permutant also corroborate the idea that the two subdomains unfold sequentially. Combining the I27A mutation with the circular permutation, both of which specifically destabilize the N-terminal subdomain and increase the curvature of their chevrons' unfolding limbs, leads to a chevron with no unfolding limb curvature. While this lack of curvature could imply that the intermediate is no longer formed, the low kinetic *m*-value of CP13*I27A (2.0 kcal/mol·M, lower than even LCTF) leads us to believe that this variant specifically destabilizes N relative to the intermediate. This mutant, like LCTF, appears to model the partially folded species seen in NSHX. On the other hand, L7A in the wild type context destabilizes the C-terminal domain to such an extent that this intermediate is no longer populated. However, when combined with CP13*, the N-terminal subdomain is also destabilized, such that J is once again populated during unfolding, as seen by the curvature in the unfolding limb of its chevron.

2.4.1 Nature of the roll-over and burst phase in the folding process

Burst phases are seen as hallmarks of intermediate formation, both during refolding and unfolding (41-44). Additionally, if representative of rapid formation of a true intermediate, they should show a cooperative, sigmoidal denaturant dependence (45). While all of the proteins and

variants investigated here show burst phase amplitudes during refolding, their values show a linear, not sigmoidal, dependence on urea concentration (Figure 3B). There are also some inconsistencies between the HX protection noted by Lu and Dahlquist in their pulse-labeling study (40) and the stability of I determined from our kinetic modeling (K_{UI}). All of the variants we have studied imply an early intermediate with a stability on the order of 4 kcal/mol, while the protection factors suggest a much lower stability (~ 2- 3 kcal/mol). The basis for the difference is unclear. Nonetheless, our results certainly argue against a specific compact structure involving residues from both subdomains as it appears insensitive to mutation and/or the complete removal of the N-terminal subdomain such as in LCTF.

2.5 Conclusions

In sum, by examining the denaturant dependence of both folding and unfolding for a diverse array of variants of T4L*, we find evidence for an unfolding intermediate that seems to be structured in only the C-terminal subdomain. This intermediate resembles the species seen by NSHX (14) and is evidence that the NSHX results from T4 lysozyme report on a species on the folded side of the rate-limiting barrier, rather than the unfolded side. NSHX detection of such species have been interpreted similarly by Bai and coworkers (46) and called "hidden intermediates" due to the inability to detect them using kinetic methods. T4 lysozyme , however, appears to be a system in which these "hidden" intermediates are actually revealed by kinetic methods. Such high energy intermediates on the folded side of the transition state barrier are likely to be as important as more commonly identified folding intermediates, if not more so, in the myriad biological functions mentioned earlier. The identification of such an intermediate species in T4L* via kinetic methods is an exciting step towards a more complete understanding of the role of intermediates on both sides of the transition state barrier in protein folding.

2.6 References

1. Privalov PL (1996) Intermediate states in protein folding. *J Mol Biol* 258(5):707-725.
2. Galzitskaya OV, Skoogarev AV, Ivankov DN, & Finkelstein AV (2000) Folding nuclei in 3D protein structures. *Pac Symp Biocomput*:131-142.
3. Englander SW (2000) Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* 29:213-238.
4. Kern D & Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13(6):748-757.
5. Busenlehner LS & Armstrong RN (2005) Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry. *Arch Biochem Biophys* 433(1):34-46.
6. Schnell JR, Dyson HJ, & Wright PE (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct* 33:119-140.
7. Bae E & Phillips GN, Jr. (2006) Roles of static and dynamic domains in stability and catalysis of adenylate kinase. *Proc Natl Acad Sci U S A* 103(7):2132-2137.
8. Voos W & Rottgers K (2002) Molecular chaperones as essential mediators of mitochondrial biogenesis. *Biochim Biophys Acta* 1592(1):51-62.
9. Veprintsev DB*, et al.* (2006) Core domain interactions in full-length p53 in solution. *Proc Natl Acad Sci U S A* 103(7):2115-2119.
10. Fink AL (2005) Natively unfolded proteins. *Curr Opin Struct Biol* 15(1):35-41.
11. Jahn TR & Radford SE (2005) The Yin and Yang of protein folding. *FEBS J* 272(23):5962-5970.
12. Stefani M & Dobson CM (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* 81(11):678-699.
13. Matsumura M & Matthews BW (1989) Control of enzyme activity by an engineered disulfide bond. *Science* 243(4892):792-794.
14. Llinas M, Gillespie B, Dahlquist FW, & Marqusee S (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* 6(11):1072-1078.
15. Cellitti J*, et al.* (2007) Exploring subdomain cooperativity in T4 lysozyme I: structural and energetic studies of a circular permutant and protein fragment. *Protein Sci* 16(5):842-851.
16. Parker MJ & Marqusee S (1999) The cooperativity of burst phase reactions explored. *J Mol Biol* 293(5):1195-1210.
17. Cantor CR & Schimmel PR (1980) *Biophysical Chemistry : Part III: The Behavior of Biological Macromolecules* (W. H. Freeman).
18. Chen BL, Baase WA, & Schellman JA (1989) Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry* 28(2):691-699.
19. Chen BL, Baase WA, Nicholson H, & Schellman JA (1992) Folding kinetics of T4 lysozyme and nine mutants at 12 degrees C. *Biochemistry* 31(5):1464-1476.
20. Desmadril M & Yon JM (1984) Evidence for intermediates during unfolding and refolding of a two-domain protein, phage T4 lysozyme: equilibrium and kinetic studies. *Biochemistry* 23(1):11-19.
21. Gassner NC*, et al.* (2003) Multiple methionine substitutions are tolerated in T4 lysozyme and have coupled effects on folding and stability. *Biophys Chem* 100(1-3):325-340.

22. Llinas M & Marqusee S (1998) Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Sci* 7(1):96-104.
23. Baldwin RL (1996) On-pathway versus off-pathway folding intermediates. *Fold Des* 1(1):R1-8.
24. Oliveberg M (2001) Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis. *Curr Opin Struct Biol* 11(1):94-100.
25. Sanchez IE & 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(4):867-884.
26. Sanchez IE & Kiefhaber T (2003) Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys Chem* 100(1-3):397-407.
27. Sanchez IE & Kiefhaber T (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* 325(2):367-376.
28. Scott KA & Clarke J (2005) Spectrin R16: broad energy barrier or sequential transition states? *Protein Sci* 14(6):1617-1629.
29. Matouschek A, Otzen DE, Itzhaki LS, Jackson SE, & Fersht AR (1995) Movement of the position of the transition state in protein folding. *Biochemistry* 34(41):13656-13662.
30. Oliveberg M, Tan YJ, Silow M, & Fersht AR (1998) The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J Mol Biol* 277(4):933-943.
31. Otzen DE, Kristensen O, Proctor M, & Oliveberg M (1999) Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry* 38(20):6499-6511.
32. Dalby PA, Oliveberg M, & Fersht AR (1998) Folding intermediates of wild-type and mutants of barnase. I. Use of phi-value analysis and m-values to probe the cooperative nature of the folding pre-equilibrium. *J Mol Biol* 276(3):625-646.
33. Bachmann A & Kiefhaber T (2001) Apparent two-state tendamistat folding is a sequential process along a defined route. *J Mol Biol* 306(2):375-386.
34. Wagner C & Kiefhaber T (1999) Intermediates can accelerate protein folding. *Proc Natl Acad Sci U S A* 96(12):6716-6721.
35. Kiefhaber T, Kohler HH, & Schmid FX (1992) Kinetic coupling between protein folding and prolyl isomerization. I. Theoretical models. *J Mol Biol* 224(1):217-229.
36. Scott KA, Randles LG, & Clarke J (2004) The folding of spectrin domains II: phi-value analysis of R16. *J Mol Biol* 344(1):207-221.
37. Scott KA, Batey S, Hooton KA, & Clarke J (2004) The folding of spectrin domains I: wild-type domains have the same stability but very different kinetic properties. *J Mol Biol* 344(1):195-205.
38. Sauder JM, MacKenzie NE, & Roder H (1996) Kinetic mechanism of folding and unfolding of Rhodobacter capsulatus cytochrome c2. *Biochemistry* 35(51):16852-16862.
39. Bieri O*, et al.* (1999) The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc Natl Acad Sci U S A* 96(17):9597-9601.
40. Lu J & Dahlquist FW (1992) Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry* 31(20):4749-4756.

41. Kuwajima K, Yamaya H, Miwa S, Sugai S, & Nagamura T (1987) Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. *FEBS Lett* 221(1):115-118.
42. Fujiwara K*, et al.* (1999) Folding-unfolding equilibrium and kinetics of equine beta-lactoglobulin: equivalence between the equilibrium molten globule state and a burst-phase folding intermediate. *Biochemistry* 38(14):4455-4463.
43. Jamin M, Yeh SR, Rousseau DL, & Baldwin RL (1999) Submillisecond unfolding kinetics of apomyoglobin and its pH 4 intermediate. *J Mol Biol* 292(3):731-740.
44. Bhuyan AK & Udgaonkar JB (1998) Multiple kinetic intermediates accumulate during the unfolding of horse cytochrome c in the oxidized state. *Biochemistry* 37(25):9147-9155.
45. Raschke TM & 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(4):298-304.
46. 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(4):757-764.

Chapter 3

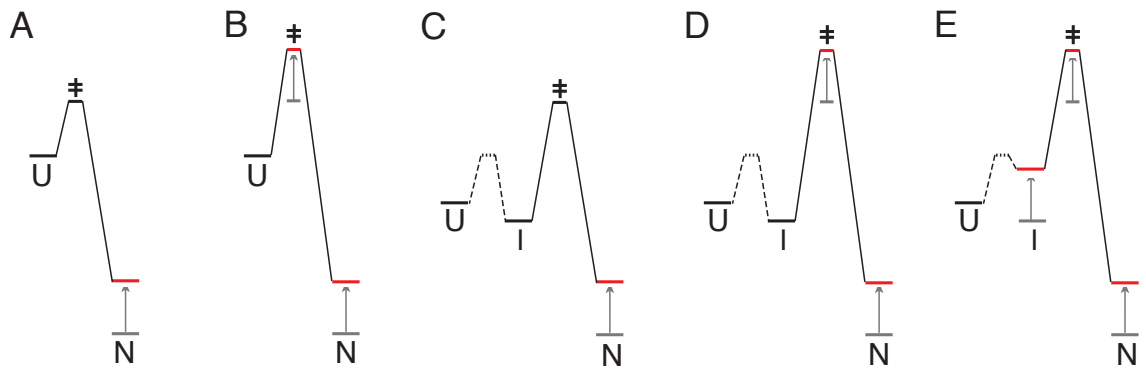
Investigating the rate-limiting barrier to folding for *E. coli* RNase H

3.1 Introduction

Despite the vast conformational space available to an unfolded protein, most are able to achieve their native conformation remarkably quickly. There is growing evidence that such efficient folding may be aided by transiently populating partially structured intermediates (1-7), which are being found in an increasing number of systems, including many that had been previously determined to be two-state (see, for example, (3, 8-17)).The existence of such species, however, indicates that while they may be productive species on the folding trajectory, they are lacking some crucial structure or interactions that must be formed in order to cross the rate-limiting transition state. The nature of the rate-limiting transition state following such intermediates, however, is still a subject of debate.

Transition-state structures have been investigated in detail in a number of mainly small, simple systems by using a Φ-value analysis approach (18), in which the kinetic effects of site-specific mutations are used to infer the level of structure with site-specific resolution in short-lived species, including the transition state and transiently populated intermediates (Figure 1). This method has been used to infer the structure of the transition state for a number of proteins, including barnase (19), various SH3 domains (20-24), chymotrypsin inhibitor 2 (25), and others (26-37); the suggested structures range from highly polarized (21, 31, 32), with some interactions essentially fully formed and others absent, to states with more diffuse structure (26, 33), where many interactions are partially formed.

Figure 1. Model reaction coordinates showing different F-value scenarios for two-state (A, B) and three-state proteins (C-E); the barrier between U and I is shown in dotted lines because it is generally not measured. (A) Destabilizing the native state but not the transition state yields a $\Phi^{\ddagger}$ of 0. (B) Destabilizing the native state to the same extent as the transition state yields a $\Phi^{\ddagger}$ of 1. (C) Destabilizing the native state but not the intermediate or transition state yields a Φ^{I} and $\Phi^{\ddagger}$ of 0. (D) Destabilizing the native state and the transition state to the same degree but leaving the intermediate unaffected yields a Φ^{I} of 0 and $\Phi^{\ddagger}$ of 1. (E) Destabilizing the native state, transition state, and intermediate to the same degree yields a Φ^{I} and $\Phi^{\ddagger}$ of 1.



Few of these studies, however, monitor the formation of structure through an intermediate species and the transition state. Such studies could begin to define the interactions that have not yet been formed or stabilized in early folding intermediates, potentially elucidating the source of the rate-limiting bottleneck. In Φ-value studies of barnase (19, 38), barstar (36), and the Fyn SH3 domain (37), all of which fold via kinetic intermediates, the majority of probes that are structured

in the transition state (high $\Phi^{\ddagger}$) are also structured to some extent in the intermediate (moderate to high Φ^{I}) implying that crossing the rate-limiting barrier mainly requires solidifying the interactions in the transition state rather than forming entirely new interactions. In contrast, studies of Im7 reveal probes that appear more structured in the intermediate than the rate-limiting transition state (Φ^{I} is higher than $\Phi^{\ddagger}$), but this unusual behavior is attributed to non-native interactions that must be broken to form the native conformation (33). There are few examples, however, of Φ-value analysis detecting specific interactions that are absent in an intermediate but completely formed upon reaching the transition state. Defining such interactions could provide powerful insight into the nature of both intermediates and post-intermediate transition states.

E. coli ribonuclease H (RNase H) provides a good model system in which to characterize such behavior. RNase H is known to fold through a productive, on-pathway, and obligatory kinetic folding intermediate (7, 29, 39-45). The precise structured regions of the intermediate remain a subject of debate, but a simple schematic model based on hydrogen exchange studies is shown in Figure 2; the region in blue – the core – is believed to be structured in the intermediate, and the remainder of the protein, or the periphery, is believed to be unstructured (39, 40). Additional support for this model has come from a limited Φ-value analysis conducted on a relatively small subset of probes either entirely buried in the core or in the periphery pointed away from the core. These studies showed that mutations in the core have high values for both Φ^{I} and $\Phi^{\ddagger}$, while mutations in the periphery have low values for both (29, 45) (Figure 2).

While much is known about the intermediate, its conversion to the fully folded state has not been well studied. It is clear that the periphery must fold and come into contact with the core, but the exact order of these events and the driving force behind them is unknown. Identifying positions that are completely unstructured in the intermediate but have native-like structure in the transition state – i.e. positions with low Φ^{I} and high $\Phi^{\ddagger}$ – is a powerful method to map out this transition. Furthermore, elucidating the process may reveal new insight into the mechanism of folding from an intermediate that may be generalizable to many proteins that fold through similar productive species.

In this study, we use a kinetic Φ-value analysis to identify positions that gain structure after the intermediate is formed, but prior to formation of the transition state. By thoroughly probing the interface between the protein core and periphery, we determine the structure of the rate-limiting transition state and identify specific contacts formed in that state that are absent in the intermediate. In doing so, we gain a deeper understanding of the nature of the rate-limiting barrier and the structural interactions required to overcome it.

3.2 Methods

3.2.1 Construction and purification of variants

All mutants were constructed by QuikChange mutagenesis. A plasmid encoding the *E. coli* RNase H* gene (* indicates that the three native cysteines have been mutated to alanine) (46) was used as the template. Most of the variants expressed solubly and were purified as described previously (46); those that expressed insolubly (V5G, T9G, A24G, and F35A) were extracted overnight in 50% acetonitrile, 0.2% TFA and dialyzed into aqueous buffer (20 mM sodium acetate pH 5.5, 50 mM KCl).

3.2.2 Equilibrium urea denaturation
Samples containing 50 ug/mL protein and varying concentrations of urea were equilibrated for the required time and the CD signal at 222 nm was measured in a 1 cm path length cuvette at 25°C using an Aviv 410 CD spectropolarimeter. Melts were repeated at least twice. The data were fit to a two-state model with a linear free-energy extrapolation.

3.2.3 Folding and unfolding kinetics
Refolding of mutants was initiated either by stopped-flow or manual mixing. For stopped-flow mixing, concentrated unfolded protein (about 8-10 mg/mL) in high urea was diluted 1:11 into final refolding conditions. The CD signal at 222 nm was monitored in an Aviv 202-SF stopped-flow CD device with a 1 mm path length cuvette; this instrument has a dead time of about 12 msec. Refolding traces were repeated at least seven times and averaged. For manual mixing, 1.5 mg/mL unfolded protein in high urea was diluted 1:30 into the indicated final refolding conditions. The CD signal at 222 nm was monitored in an Aviv 410 CD using a 1 cm path length cuvette; dead times ranged from about 5 to 10 seconds.

Unfolding kinetics were also measured by manual mixing using a similar protocol. Folded protein (1.5 mg/mL in low urea) was diluted 1:30 into final unfolding conditions. The CD signal at 222 nm was monitored in an Aviv 410 CD using a 1 cm path length cuvette; dead times ranged from about 5 to 10 seconds. All kinetic data were adequately fit by a three-parameter single exponential and used to construct chevron plots for each variant. Rates, burst phase amplitudes, and final amplitudes were simultaneously fit to a three-state on-pathway intermediate model (7).

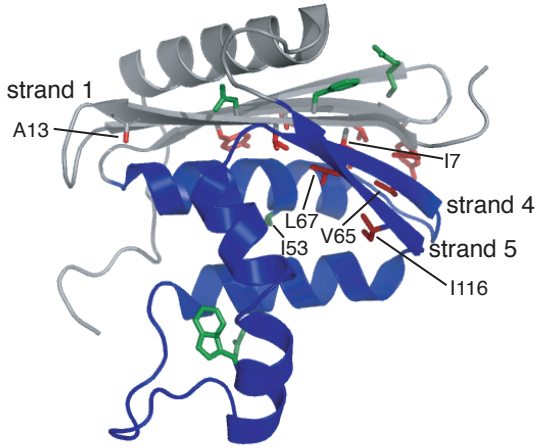
3.2.4 Calculation of Φ-values
Calculations of both Φ^I and $\Phi^{\ddagger}$ depend upon the $\Delta\Delta G_{UN}$ determined for the mutant versus the wild-type. These values were calculated two different ways – based on equilibrium denaturation, and based on the total stability as calculated from the fit of the chevron – resulting in two values each for Φ^I and $\Phi^{\ddagger}$.

$$\Phi^I = -RT \ln(K_{UI(wt)}/K_{UI(mut)})/(\Delta G_{UN(wt)} - \Delta G_{UN(mut)})$$

$$\Phi^{\ddagger} = 1 - (RT \ln (k_{ni(mut)}/k_{ni(wt)})/(\Delta G_{UN(wt)} - \Delta G_{UN(mut)}))$$

3.3 Results
A number of point mutants were designed in the background of *E. coli* RNase H* (the asterisk denotes a version of the protein in which the three native cysteines have been mutated to alanine (46)). Mutants were chosen that were sufficiently destabilizing for a meaningful interpretation of the kinetic effect ($\Delta\Delta G \geq$ about 1.0 kcal/mol). The majority of the positions mutated were hydrophobic groups such as valine, isoleucine, leucine, and alanine. In most cases, these side chains were replaced by alanine; in cases where such a mutation was not appropriately destabilizing or the residue began as an alanine, other residues were introduced. Mutated residues reside on the β sheet with side chains that face toward the protein core. Figure 2 shows a structure of the protein, highlighting the residues mutated for this study in red; residues in green have been studied in previous Φ-value analysis of this protein.

Figure 2. *E. coli* RNase H*; the core region believed to be structured in the folding intermediate is shown in blue, and side chains mutated for Φ-value analysis are shown in sticks. Those in green were investigated in previous studies; those in red are described in this chapter.



Equilibrium stabilities of all mutants were measured by circular dichroism (CD) measurement of urea-induced denaturation; results are reported in Table 1. Folding and unfolding kinetics were measured by manual mixing and stopped-flow CD. In all cases, the resulting chevron plots revealed a linear unfolding limb and curvature in the folding limb (Figure 3), similar to the kinetics reported for the wild-type protein (29). All chevron plots were fit to a three-state on-pathway model with the roll-over in the folding limb attributed to rapid accumulation (within the burst phase of the instrument) of a transient intermediate at low denaturant concentrations (40). In some of the variants investigated here, the roll-over was subtle; however, all variants showed significant burst phase amplitudes attributed to the formation of the intermediate (Figure 4), confirming the validity of the three-state model.

The variants can be clearly divided into three classes based on their kinetic behavior: those with little to no effect on the unfolding limb but a large effect on the folding limb (Group 1; Figure 3A), those with a moderate effect on both the folding and unfolding limbs (Group 2; Figure 3C,D), and those with an effect on the unfolding limb but little to no effect on the folding limb (Group 3; Figure 3E). One variant, I7A, greatly slowed both the folding and unfolding rates, but additional mutations at this position had essentially no effect on the unfolding limb (Figure 3B), resulting in its classification in Group 1. These groupings appear to define specific structural regions (Figure 3F), suggesting regions that are fully, moderately, or completely lacking structure in the transition state (Groups 1, 2, and 3, respectively).

The chevron fits were also used to calculate Φ^{I} and $\Phi^{\ddagger}$, listed in Table 1 (see Materials and Methods), and the classification system described previously is supported by the Φ-values. Values of $\Phi^{\ddagger}$ for Group 1 probes (excluding I7A due to its unusual behavior) range from 0.84 to 1.29, suggesting that these positions are completely structured in the transition state. Group 2 variants have $\Phi^{\ddagger}$-values from 0.44 to 0.89, indicating partial structure in the transition state, and A13G, the sole variant in Group 3, has a $\Phi^{\ddagger}$ of approximately zero, suggesting a complete lack of structure. Φ^{I}-values, on the other hand, do not track with the groupings – as expected. Probes

from all groups have relatively low Φ^I-values (-0.49 to 0.35); only two variants, L67A and I116A, have moderate Φ^I-values, of about 0.5 and 0.6, respectively. Both of these probes are in Group 3.

Figure 3. Chevron plots for β sheet variants; the wild-type chevron is shown in a dashed line for comparison. (A) Chevron plots for mutations at positions that have little to no effect on the unfolding limb (Group 1). I7V, filled circles; V65G, filled squares; L67A, open circles. (B) Chevron plots for multiple mutations at the I7 position. I7A, filled circles; I7V, filled squares; I7L, open circles. (C and D)Chevron plots for mutants that have moderate effects on both the folding and unfolding limbs (Group 2), divided into two groups for clarity. (C)V5G, filled circles; T9G, open squares; Y22V, filled squares; A24G, open circles. (D) L26A, filled squares; Y28A, filled circles; F35A, open squares; I116A, open circles. (E) Chevron plot for A13G (Group 3), which has almost no effect on the folding limb. (F) A structure of the protein with the side chains investigated shown in sticks; Group 1 is red, Group 2 orange, Group 3 yellow.

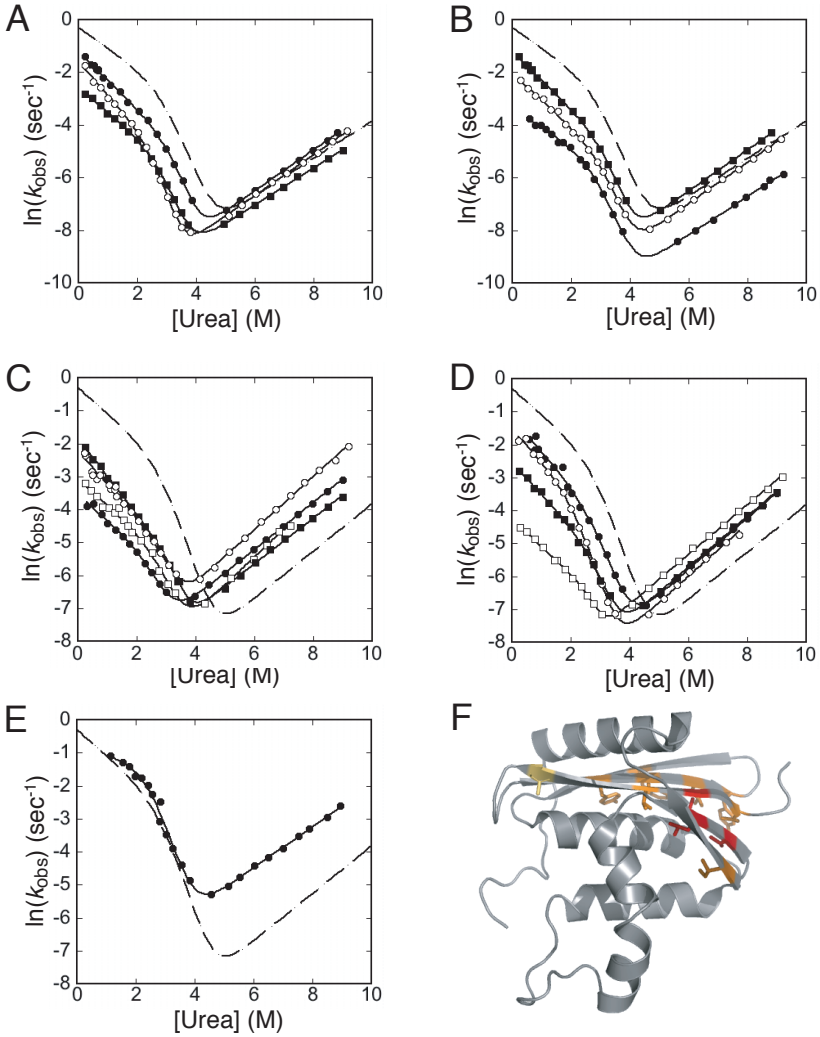
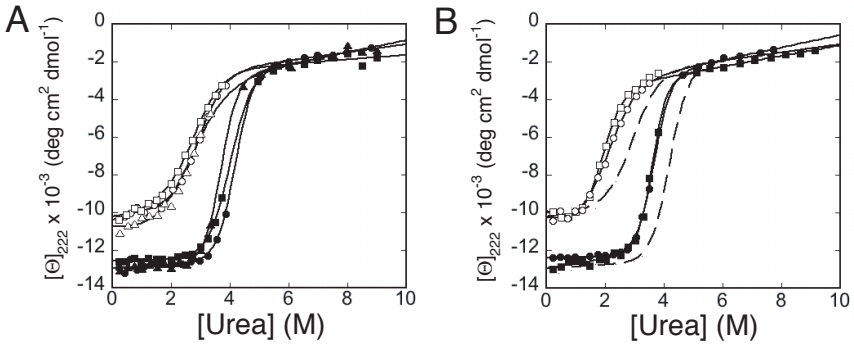


Table 1. Summary of the equilibrium and kinetic fits for all mutants. Reported Φ-values are based on ΔΔG as measured by equilibrium denaturation. Free energies are in units of kcal/mol; *m*-values are in units of kcal/mol M^{-1}; rate constants are in units of sec^{-1}.

	$\Delta G°_{urea\ melt}$	$m_{urea\ melt}$	$\Delta G°_{kinetic}$	$m_{kinetic}$	K_{ui}	m_{ui}	$k_{in}(H_20)$	m_{in}	$k_{ni}(H_20)$	m_{ni}	Φ^I	$\Phi^‡$
RNH[a]	9.7 ± 0.4	2.1 ± 0.1	9.9	2.1	400	1.24	0.74	0.45	1.7E-5	0.42		
V5G	6.6 ± 0.3	2.0 ± 0.1	6.8	2.1	279	1.25	0.029	0.48	8.1E-5	0.42	0.07	0.70
I7A	8.6 ± 0.3	2.1 ± 0.1	9.4	2.3	875	1.36	0.039	0.49	4.0E-6	0.42	-0.43	1.78
I7V	8.7 ± 0.2	2.1 ± 0.1	9.5	2.3	672	1.33	0.26	0.56	1.8E-5	0.44	-0.32	0.96
I7L	8.9 ± 0.2	2.1 ± 0.1	9.0	2.2	410	1.29	0.11	0.49	1.1E-5	0.44	-0.02	1.29
T9G	6.3 ± 0.5	1.8 ± 0.1	6.8	1.9	65	0.94	0.046	0.43	3.2E-5	0.48	0.31	0.89
A13G	8.6 ± 0.3	2.1 ± 0.1	9.8	2.5	3491	1.52	1.37	0.60	3.0E-4	0.37	-1.22	-0.62
Y22V	8.3 ± 0.2	2.1 ± 0.1	8.6	2.2	341	1.23	0.33	0.55	5.8E-5	0.41	0.07	0.46
A24G	7.3 ± 0.2	2.0 ± 0.1	7.5	2.2	301	1.25	0.098	0.51	8.8E-5	0.47	0.07	0.60
L26A	7.3 ± 0.3	2.0 ± 0.1	9.2	2.7	2943	1.70	0.071	0.53	4.1E-5	0.44	-0.49	0.79
Y28A	8.5 ± 0.4	2.1 ± 0.1	8.4	2.2	220	1.20	0.26	0.51	3.6E-5	0.44	0.29	0.63
F35A	6.0± 0.3	1.8 ± 0.1	6.2	2.0	152	1.17	0.013	0.44	5.4E-5	0.44	0.15	0.82
V65G	8.1 ± 0.2	2.1 ± 0.1	8.6	2.3	434	1.32	0.074	0.53	1.6E-5	0.40	-0.03	1.03
L67A	8.1 ± 0.2	2.2 ± 0.1	8.1	2.3	70	1.37	0.15	0.43	1.1E-5	0.47	0.66	1.16
I116A	8.3 ± 0.2	2.3 ± 0.1	8.0	2.2	86	1.27	0.21	0.51	2.5E-5	0.45	0.64	0.84

[a] These values were previously determined by Raschke *et al*. (29)

Figure 4. Final signals (filled symbols) and burst phase signals (open symbols) for a representative sample of variants. All variants had burst phase amplitudes that showed sigmoidal dependence on [Urea] (A) The burst phase amplitudes for a representative sample of variants with burst phase amplitudes that show no destabilization of the intermediate. I7V, circles; L26A, triangles; V65G, squares. (B) Burst phases for the two cases in which the intermediate is slightly destabilized. L67A, squares; I116A, circles; the fit for I7V is shown in dotted lines as a reference.



3.4 Discussion

3.4.1 Refining the structural model for the kinetic intermediate

The structural interpretation of the kinetic folding intermediate for *E. coli* RNase H has arisen from a number of different sources. Pulsed hydrogen exchange experiments suggest that structured regions for this species include helices A and D, as well as strand 4 (40). Equilibrium hydrogen exchange experiments on the acid-induced molten globule state of the protein, which is believed to potentially reflect the structure of the early folding intermediate, show major protection for helices A and D, with strand 4 protected to a lesser extent, though more than most of the other strands (47). Native-state hydrogen exchange experiments on the protein, as well as a hyperstabilized variant, detect a conformation present at equilibrium with helices A and D and strand 4 protected, although in both of these cases the protection for strand 4 is slightly lower

than that for the helices (39, 48). The similarity between this structure and the structure seen in previous hydrogen exchange experiments led to the belief that the two are correlated.

In an attempt to model the folding intermediate, a fragment containing helices A through D and strand 4 was created (41). This fragment shows cooperative, apparently two-state folding and secondary structure similar to the core of full-length RNase H, but it was found to exist in a monomer-dimer equilibrium, complicating further biophysical characterization. It is possible, however, that this fragment nevertheless represents the folding intermediate and the unstructured β-sheet region that would be present in the full-length protein plays an important role in blocking dimerization.

Computational studies expanded this picture of the folding intermediate. The RAFT (rapid autonomous folding test) computation suggested that helices A through D, along with strands 4 and 5, form an "autonomous folding unit", which may reflect the structure of the folding intermediate (42). This was the first time that strand 5 was implicated in the folding intermediate; in previous hydrogen exchange experiments, probes in this region were rare, and those that were measured did not show greater protection than that seen for strands 1-3. Thus, it remained unclear to what extent strands 4 and 5 should be included in the structural model of the folding intermediate.

The kinetic analysis carried out here provides a somewhat complicated answer to this question. Based on burst phase amplitudes and Φ^I, it appears that L67A (strand 4) and I116A (strand 5) slightly destabilize the intermediate (Figure 4); however, V65G (strand 4) does not. The Φ^I-values for L67A and I116A are quite similar but moderate (0.66 and 0.64, respectively), suggesting that these positions are only partially involved in their native-like energetic interactions in the intermediate. Though V65G does not show destabilization of the intermediate, it seems unlikely that it would have significantly different structure than L67 in the intermediate, and HX has strongly implicated that this strand is structured in the intermediate. The lack of effect on the intermediate upon mutation at V65 can be explained by its location further from the traditional core region relative to L67; while the V65G mutation may substantially destabilize the native state upon complete packing, in the intermediate state, it is more of a surface position and therefore more tolerant to mutation.

The results for I116A suggest that structure in strand 5 in the intermediate may have been missed in previous HX experiments. Strand 5 is at the edge of the β sheet, so its lack of protection by HX could be attributed to a lack of hydrogen bonding partners and possible backbone fluctuations despite well-packed side-chains. In this case, investigating side-chain packing by Φ-value analysis may be more structurally instructive.

Thus, a combination of backbone amide hydrogen exchange experiments and side chain-based Φ-value analysis would suggest that the intermediate may be most accurately modeled by including helices A through D and strands 4 and 5, as indicated by the RAFT calculation and reflected in Figure 2. A recent study on *Thermus thermophilus* RNase H, a homolog with about 60% similarity to the *E. coli* protein, modeled the intermediate by including strand 1 (49) based on a single hydrogen exchange pulse-labeling probe in the strand that shows partial protection

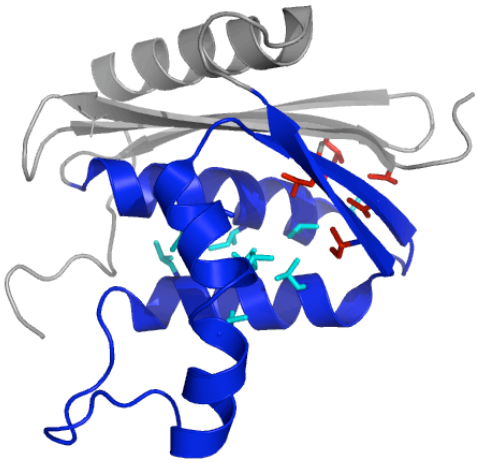
(50). The Φ-values measured here, however, suggest that, at least in the *E. coli* protein, strand 1 is not involved in the kinetic intermediate.

3.4.2 A nucleation core for transition state formation

The experiments conducted here suggest that three side chains on strands 1 and 4 – I7, V65, and L67 (Group 3) – act as a nucleation core for transition state formation. Radiating out from this crucial core, side chains appear to have a diminishing effect on transition state formation, reaching the extreme case of A13G, which lies on the opposite end of strand 1 from I7 and has essentially no effect. The significant deviation in behavior between these two probes on the same element of secondary structure indicates that the β sheet does not form its final structure in one concerted step, but in a more gradual process that begins at the nucleation core.

There are several other lines of evidence that support this model of a post-intermediate nucleation core. Native-state thiol alkyl-proton exchange experiments on a hyperstabilized mutant of *E. coli* RNase H indicated that I7 (strand 1) and A24 (strand 2) become structured before the rate-limiting barrier in the folding pathway (RB *et al.*, submitted; Chapter 5); such behavior was not seen for other positions in the strands further from the hypothesized nucleation core. Computational ILV cluster analysis, which identifies clusters of contacting isoleucine, leucine, and valine residues that have been implicated as an important driving force for early folding events (51), also highlights the importance of the region investigated in this study (S. Kathuria, personal communication). *E. coli* RNase H contains a single ILV cluster that spans both the core and the periphery of the protein (Figure 5); those that reside in the core (blue) are distributed across the region relatively evenly, but those in the periphery are limited to the region around residues 7 and nicely mirror the nucleation core suggested by the positions with high $\Phi^{\ddagger}$ values in this study. Thus it appears that, in this case, the ILV cluster may determine not only collapse to the intermediate, but also the next step of nucleation to cross the rate-limiting transition state.

Figure 5. ILV cluster residues shown as sticks; intermediate structure as defined by previous experiments shown in blue. For clarity, ILV cluster residues in the core helices are shown in cyan; those in the strands are shown in red.



These results support a nucleation-condensation model, in which a hydrophobic core acts as a template upon which secondary and tertiary structure can form. Previously, however, this model has predominately been applied to the formation of early intermediates or the transition states in cases where intermediates are not detected (see, for example, (25, 31, 36, 52, 53)). In the case of RNase H, it appears that there may be two subsequent nucleation-condensation processes: the formation of the intermediate, and a subsequent nucleation to reach the post-intermediate transition state.

Based on a model of the intermediate in which strand 4 is structured, the Φ-values for probes in this strand suggests that there is a crucial conformational switch that must occur between populating the intermediate and reaching the transition state. Specifically, it appears that though this strand, as determined largely by backbone HX measurements, is structured in the intermediate, the side chains have not yet taken on specific sterically packed conformations, and perhaps it is only upon reaching such a properly packed state that the protein can traverse the rate-limiting barrier. Such behavior would result in the low Φ^{I} but high $\Phi^{\ddagger}$ for the relatively conservative hydrophobic mutations applied in this investigation, suggesting that the rate-limiting transition state from the intermediate to the native state is largely dependent upon precise steric packing in addition to hydrophobic nucleation, in contrast to the established nucleation-condensation model for early folding.

3.4.3 I7 as a lynchpin for transition state formation

Isoleucine 7 stands within the transition state nucleation core for the unusually pronounced effect upon introducing mutations at this position. The I7A mutation has a particularly dramatic effect, slowing both folding and unfolding and resulting in a $\Phi^{\ddagger}$ of greater than one. There are a number of possible explanations for Φ-values greater than one, including non-native interactions. In this case, however, because other mutations at this position do not significantly perturb the unfolding limb, it seems unlikely that such non-native interactions are a fundamental component of the folding pathway at this position. Instead, it appears that removing a significant amount of the side chain bulk with the substitution to alanine may disturb the crucial native contacts at this position sufficiently that new, non-native contacts are formed that slow folding. Furthermore, even the isosteric mutation to leucine destabilizes the native state and results in a $\Phi^{\ddagger}$ that is greater than one, suggesting that transition state formation requires very precise steric contacts at this position.

A similar series of mutations made at position I53, which resides in the core (7) (Figure 2). Mutating I53 to alanine and phenylalanine resulted in $\Phi^{I} \approx \Phi^{\ddagger} \approx 1$, as expected for a residue in the core, while I53L has no effect on the stability of the intermediate or the transition state ($\Phi^{I} \approx \Phi^{\ddagger} \approx 0$), implying that the slightly different packing resulting from a leucine rather than an isoleucine at that position only becomes important after crossing the rate-limiting transition state. I53V was not sufficiently destabilized for Φ-value analysis. The results of the leucine and valine mutants suggest that the core isoleucine is more robust to conservative hydrophobic mutations, while I7 is highly sensitive to very specific packing interactions. This interpretation is consistent with a model in which the intermediate, with the core region structured, is relatively molten; the I53L Φ-values suggest that this moltenness in the core region may persist until the transition state to folding is traversed (7). The role of isosteric mutations in the periphery was further

investigated at position L67, which is adjacent to I7 and forms part of the Group 1 nucleation core defined earlier. Unlike I7L, however, the L67I mutation did not result in any destabilization (data not shown), highlighting the unique structural importance of I7 in both the transition state and the native state.

3.5 Conclusions

In these experiments, we identify probes that, based on their Φ-values, are not involved in energetic interactions in the intermediate but adopt native-like contacts in the transition state. This is an important step toward determining the role of intermediates on the folding landscape and determining the bottleneck for the transition from an intermediate to a fully folded form.

Furthermore, the identification of a polarized transition state based on apparent hydrophobic side-chain nucleation suggests that, in the case of *E. coli* RNase H, it is side-chain packing rather than backbone hydrogen bonding that drives the protein from the partially folded intermediate to the fully folded native state. It is possible that this result is general and that the nature of the rate-limiting barrier from an intermediate to the folded state relies on side-chain packing based around a hydrophobic core with highly specific steric packing.

3.6 References

1. Khorasanizadeh S, Peters ID, & Roder H (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* 3(2):193-205.
2. Heidary DK, O'Neill JC, Jr., Roy M, & Jennings PA (2000) An essential intermediate in the folding of dihydrofolate reductase. *Proc Natl Acad Sci U S A* 97(11):5866-5870.
3. 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(46):16650-16655.
4. Wagner C & Kiefhaber T (1999) Intermediates can accelerate protein folding. *Proc Natl Acad Sci U S A* 96(12):6716-6721.
5. Friel CT, Beddard GS, & Radford SE (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(1):261-273.
6. Goldbeck RA, Chen E, & Kliger DS (2009) Early events, kinetic intermediates and the mechanism of protein folding in cytochrome C. *Int J Mol Sci* 10(4):1476-1499.
7. Spudich GM, Miller EJ, & 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(2):609-618.
8. Matouschek A, Kellis JT, Jr., Serrano L, Bycroft M, & Fersht AR (1990) Transient folding intermediates characterized by protein engineering. *Nature* 346(6283):440-445.
9. Bachmann A & Kiefhaber T (2001) Apparent two-state tendamistat folding is a sequential process along a defined route. *J Mol Biol* 306(2):375-386.
10. Sanchez IE & Kiefhaber T (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* 325(2):367-376.
11. Teilum K, Maki K, Kragelund BB, Poulsen FM, & 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(15):9807-9812.
12. Ferguson N, Capaldi AP, James R, Kleanthous C, & Radford SE (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* 286(5):1597-1608.
13. Jemth P, *et al*. (2004) Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc Natl Acad Sci U S A* 101(17):6450-6455.
14. 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(4):757-764.
15. Feng H, Vu ND, & Bai Y (2005) Detection of a hidden folding intermediate of the third domain of PDZ. *J Mol Biol* 346(1):345-353.
16. Korzhnev DM, *et al*. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430(6999):586-590.
17. Khorasanizadeh S, Peters ID, Butt TR, & Roder H (1993) Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* 32(27):7054-7063.
18. Matouschek A, Kellis JT, Jr., Serrano L, & Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340(6229):122-126.

19. Serrano L, Matouschek A, & Fersht AR (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(3):805-818.
20. Martinez JC, Pisabarro MT, & Serrano L (1998) Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol* 5(8):721-729.
21. Grantcharova VP, Riddle DS, Santiago JV, & Baker D (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* 5(8):714-720.
22. Fulton KF, Main ER, Daggett V, & Jackson SE (1999) Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J Mol Biol* 291(2):445-461.
23. Northey JG, Di Nardo AA, & Davidson AR (2002) Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol* 9(2):126-130.
24. Zarrine-Afsar A, Lin SL, & Neudecker P (2010) Mutational investigation of protein folding transition states by Phi-value analysis and beyond: lessons from SH3 domain folding. *Biochem Cell Biol* 88(2):231-238.
25. Itzhaki LS, Otzen DE, & Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254(2):260-288.
26. Bueno M, Ayuso-Tejedor S, & Sancho J (2006) Do proteins with similar folds have similar transition state structures? A diffuse transition state of the 169 residue apoflavodoxin. *J Mol Biol* 359(3):813-824.
27. Curnow P & Booth PJ (2009) The transition state for integral membrane protein folding. *Proc Natl Acad Sci U S A* 106(3):773-778.
28. Huysmans GH, Baldwin SA, Brockwell DJ, & Radford SE (2010) The transition state for folding of an outer membrane protein. *Proc Natl Acad Sci U S A* 107(9):4099-4104.
29. Raschke TM, Kho J, & Marqusee S (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* 6(9):825-831.
30. Calosci N, *et al*. (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc Natl Acad Sci U S A* 105(49):19241-19246.
31. Went HM & Jackson SE (2005) Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* 18(5):229-237.
32. Garcia-Mira MM, Boehringer D, & Schmid FX (2004) The folding transition state of the cold shock protein is strongly polarized. *J Mol Biol* 339(3):555-569.
33. Capaldi AP, Kleanthous C, & Radford SE (2002) Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* 9(3):209-216.
34. Bartlett AI & Radford SE (2010) Desolvation and development of specific hydrophobic core packing during Im7 folding. *J Mol Biol* 396(5):1329-1345.
35. Anil B, Sato S, Cho JH, & Raleigh DP (2005) Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing. *J Mol Biol* 354(3):693-705.
36. Nolting B, *et al*. (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci U S A* 94(3):826-830.
37. Neudecker P, Zarrine-Afsar A, Davidson AR, & Kay LE (2007) Phi-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc Natl Acad Sci U S A* 104(40):15717-15722.

38. Matouschek A, Serrano L, & Fersht AR (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(3):819-835.
39. Chamberlain AK, Handel TM, & Marqusee S (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* 3(9):782-787.
40. Raschke TM & 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(4):298-304.
41. Chamberlain AK, Fischer KF, Reardon D, Handel TM, & Marqusee AS (1999) Folding of an isolated ribonuclease H core fragment. *Protein Sci* 8(11):2251-2257.
42. Fischer KF & Marqusee S (2000) A rapid test for identification of autonomous folding units in proteins. *J Mol Biol* 302(3):701-712.
43. Cecconi C, Shank EA, Bustamante C, & Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. *Science* 309(5743):2057-2060.
44. Connell KB, Horner GA, & 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(2):461-470.
45. Connell KB, Miller EJ, & 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(2):450-460.
46. Dabora JM & Marqusee S (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: characterization of a partially folded state. *Protein Sci* 3(9):1401-1408.
47. Dabora JM, Pelton JG, & Marqusee S (1996) Structure of the acid state of Escherichia coli ribonuclease HI. *Biochemistry* 35(37):11951-11958.
48. Goedken ER & Marqusee S (2001) Native-state energetics of a thermostabilized variant of ribonuclease HI. *J Mol Biol* 314(4):863-871.
49. 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(2):531-539.
50. Hollien J & Marqusee S (2002) Comparison of the folding processes of T. thermophilus and E. coli ribonucleases H. *J Mol Biol* 316(2):327-340.
51. Wu Y, Vadrevu R, Kathuria S, Yang X, & Matthews CR (2007) A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the alpha subunit of tryptophan synthase, a TIM barrel protein. *J Mol Biol* 366(5):1624-1638.
52. Fersht AR (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* 92(24):10869-10873.
53. Nolting B & Agard DA (2008) How general is the nucleation-condensation mechanism? *Proteins* 73(3):754-764.

Chapter 4

Developing native-state thiol alkyl-proton exchange with *E. coli* RNase H

Conducted in collaboration with Kierstin L. Schmidt and Pehr B. Harbury, Stanford University

4.1 Introduction

Identification and characterization of structural fluctuations that occur under native conditions is crucial for understanding protein folding and function, but such fluctuations are often rare and transient, making them difficult to study. Native-state hydrogen exchange (NSHX) has been a powerful tool for identifying and providing site-specific structural information about such rarely populated conformations (see, for example, (1-8)), but it provides little insight into side-chain packing and thus provides an incomplete picture regarding the structures of these species.

To complement such studies, there is interest in developing native-state thiol proton exchange, a method analogous to NSHX that monitors reactions of cysteine side chains with modification agents rather than backbone amide exchange with solvent. The unique reactivity of the cysteine provides a site-specific probe that can be used to investigate side chain involvement in partially folded forms on the folding landscape (see, for example, (9-15)). Monitoring of cysteine reactivity has also been used to predict interfaces for protein-protein interactions (16), determine protein global stabilities (17), and to detect protein deformation in live cells (18).

The primary limiting factor for such studies is the requirement for cysteine at the position(s) of interest. In cases where large-scale, site-specific structural characterization analogous to that allowed by NSHX is desired, cysteines must be distributed throughout the protein, and modifications must be detected in a site-specific manner. Cysteines can be introduced by site-directed mutagenesis and modification of mutants can be monitored individually, but such a procedure becomes cumbersome for the full characterization of any but the smallest proteins.

To ease such studies, Silverman and Harbury developed a library of mutant cysteine tRNAs that incorporate cysteine in place of a desired amino acid and a gel-based system for simultaneous site-specific detection of modification at multiple thiol positions (19). This method uses a mutant tRNA-based *in vivo* misincorporation system to create a pool of single cysteine variants that are then exposed to thiol-reactive modifying agents as a function of time and denaturant concentration to monitor partial and full unfolding events under native conditions. Using this system, they characterized the native-state exchange behavior of 47 positions throughout the yeast TIM protein, revealing two partially unfolded forms (14). This protein, however, has not been characterized by other, more traditional biophysical methods; thus, it is unclear how these results relate to those obtained by traditional equilibrium studies or NSHX and it is possible that the cysteine mutations or other factors in the study unduly influenced the results. Applying this method to a more well-studied protein, however, could address some of these concerns.

E. coli ribonuclease H (RNase H; Figure 1A) has been characterized in great detail by a variety of traditional and innovative biophysical methods, including circular dichroism (CD) (20-23), hydrogen exchange (3, 22), X-ray crystallography (24), NMR (25, 26), pulsed proteolysis (27), and single molecule mechanical unfolding (28). A NSHX study on this protein revealed partially unfolded forms present under native conditions (3), suggesting that it is a promising candidate for NSSX studies, which could reveal previously undetected side-chain interactions in these conformations. Furthermore, many mutations have been characterized, including the D10A variant, which slows unfolding and stabilizes the protein (21); this variant has also been characterized by NSHX (29). The wealth of knowledge about RNase H and a number of variants

from NSHX and other studies provides a solid foundation from which to apply a novel method like NSSX.

Prior to full NSSX characterization, the kinetic exchange regime must be determined. The Linderstom-Lang model describes the exchange process (30), in which a protein in the "closed" or native conformation must undergo a fluctuation to some alternative "open" conformation in order to exchange:

$$closed \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} open \xrightarrow{k_{int}} exchanged$$

Based on this model, the observed rate of exchange for any probe can be written as $k_{int} \bullet k_{op}/(k_{int} + k_{cl})$.

Depending on the relative kinetics of these processes, the observed exchange rate can report on either the rate of forming the open state, k_{op} (EX1 regime, where $k_{cl} << k_{int}$), or on the stabilities of open forms, K_{op} (EX2 regime, where $k_{cl} >> k_{int}$) (31). An additional exchange regime termed EXX is observed when $k_{cl} \approx k_{int}$ (32); in this case, simplification of the observed rate expression is not possible and analysis is more complicated. The kinetic exchange regime for each probe can be determined experimentally by measuring the exchange rate as a function of the concentration of modifying reagent; EX1 exchange is independent of [IAM], while EX2 exchange shows a linear dependence with a slope of 1, and EXX probes exhibit intermediate slopes.

In this chapter, I describe the construction of *E. coli* RNase H variants suitable for NSSX. The equilibrium and kinetic behavior of cysteine-free versions of these variants under the conditions of the NSSX experiment were monitored by CD, and the kinetic exchange regimes for a sampling of the targeted cysteine variants were also determined. We discovered a number of problems with the pseudo-wild type protein that precluded a full NSSX characterization, but found that these issues were relieved by working in the background of a previously characterized variant.

4.2 Methods

4.2.1 Gene construction

Constructs of *E. coli* RNase H*, a variant of *E. coli* RNase H with its three free cysteine residues replaced with alanine (20), and D10A *E. coli* RNase H* (21) with N-terminal 6-His tags and C-terminal ybbR tags (33) for enzymatic labeling with a fluorophore were created. Individual cysteine mutations were introduced by QuikChange mutagenesis.

4.2.2 Protein expression and purification

All proteins were expressed in BL21(DE3) pLysS *E. coli* cells grown in LB medium. For alkylation experiments, single colonies of three to eight single-cysteine variants were grown separately overnight and LB medium was inoculated with culture from each of these variants. In all cases, expression was induced at OD_{600} = 0.6-0.8 with 1 mM IPTG and cells were harvested after 2-3 hours.

For CD experiments, cells were resuspended in 50 mM Tris pH8, 50 mM NaCl and run over a heparin column with a gradient to 500 mM NaCl. Fractions containing the protein were pooled

and subsequently loaded on a Ni-NTA column, washed with 50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole, and eluted at 500 mM imidazole. Protein was then dialyzed into 20 mM Tris pH 8.6, 50 mM KCl. For CD experiments of single-cysteine mutants, all buffers also contained 1 mM TCEP. For alkylation experiments, the heparin column was omitted. Cells were resuspended in 20 mM Tris pH 8, 20 mM imidazole, 500 mM NaCl, 1 mM TCEP and purified on a Ni-NTA column as described above. Protein was then dialyzed into 100 mM bicine pH 8.6, 50 mM KCl, 1 mM TCEP.

4.2.3 Circular dichroism

For equilibrium melts, CD data were collected using an Aviv 410 spectropolarimeter with a Peltier temperature-controlled sample holder and 1 cm pathlength cuvette. Protein denaturation studies were conducted by monitoring the ellipticity at 222 nm as a function of [GdmCl] at 25°C. Samples were allowed to equilibrate overnight, and the signal for each sample was averaged over 60 seconds. Free energies of unfolding and *m*-values were calculated using Kaleidagraph assuming a two-state transition and linear dependence on denaturant (34). Kinetic measurements were taken with an Aviv 202 spectropolarimeter with a stopped-flow system or an Aviv 410 spectropolarimeter; ellipticity at 222 nm was measured as a function of time. Experiments conducted on the 202 were repeated seven times and averaged; data on the 410 were taken once and spot-checked for reproducibility.

4.2.4 Thiol alkyl-proton exchange

To determine the kinetic exchange regime, pools of cysteine-containing variants were incubated with 1, 5, 10, and 25 mM iodoacetamide (IAM) at varying concentrations of GdmCl; alkylation was quenched at various time points with an excess of DTT. Samples were prepared using a modified scheme based on that originally used by Silverman and Harbury (19). Protein was labeled with fluorescein-CoA by the enzyme Sfp phosphopantetheinyl transferase (33), denatured in 4 M GdmCl pH 8.6, cyanylated with dinitrothiocyanobenzoic acid (NTCB), and precipitated with 0.5% trichloroacetic acid. Pellets were washed twice with acetone, resuspended in 8 M urea in 1 M ammonium hydroxide, and incubated at room temperature for 1 hr, resulting in cleavage at the positions of cyanylation. Fragments were separated by gel electrophoresis and fluorescein-coA-labeled protein was visualized with a Typhoon fluorescence imager. Band intensities were quantified using ImageQuant and normalized based on a syntaxin loading control. Normalized data were fit to a three-parameter single exponential using Kaleidagraph. See appendix for further details.

4.3 Results

4.3.1 Characterization of RNase H and cysteine variants

A version of *E. coli* RNase H* (* denotes a variant of the protein with the three native cysteines mutated to alanine (20)) suitable for our thiol exchange experiments was created by the addition of an N-terminal 6-His tag and a C-terminal ybbR tag for fluorescent labeling (33); this variant is referred to here simply as RNH.

For effective thiol reactivity, thiol exchange experiments were carried out at pH 8.6. Previous characterization of E. coli RNase H, however, had mostly been done at pH 5.5. The stabilities for both RNH* and RNH were measured under the conditions for NSSX experiments and were found to be identical (Table 1). The folding and unfolding kinetics were also determined (Figure

1B). The resulting chevron plot reveals the characteristic rollover in the folding limb attributed to the presence of a folding intermediate, as expected based on previous folding studies of *E. coli* RNase H* (22).

Stabilities of a number of single cysteine variants were also measured by CD under NSSX conditions (Table 1). Some showed small to moderate destabilization and modest *m*-value effects due to cysteine introduction. I7C and L59C, however, had extremely compromised *m*-values; in the extreme case of L56C, the transition is so broad that the data cannot be fit reliably (data not shown). A similarly low *m*-value was seen in an earlier variant of RNase H and was attributed to an increased population of the intermediate under native conditions (35). It is likely that these cysteine mutations, and perhaps others as well, have a similar affect. In this case, interpretation of the NSSX results, which relies on the native state being the predominantly populated species under the experimental conditions, could be compromised. Each cysteine mutant could be tested for this behavior to determine the pool of mutations available for NSSX; such characterization, however, was not pursued because of other difficulties with the construct, described below.

Table 1. Stabilities of *E. coli* RNase H, measured by equilibrium GdmCl melt monitored by CD. Stabilities were measured under NSSX conditions (pH 8.6), unless otherwise noted.

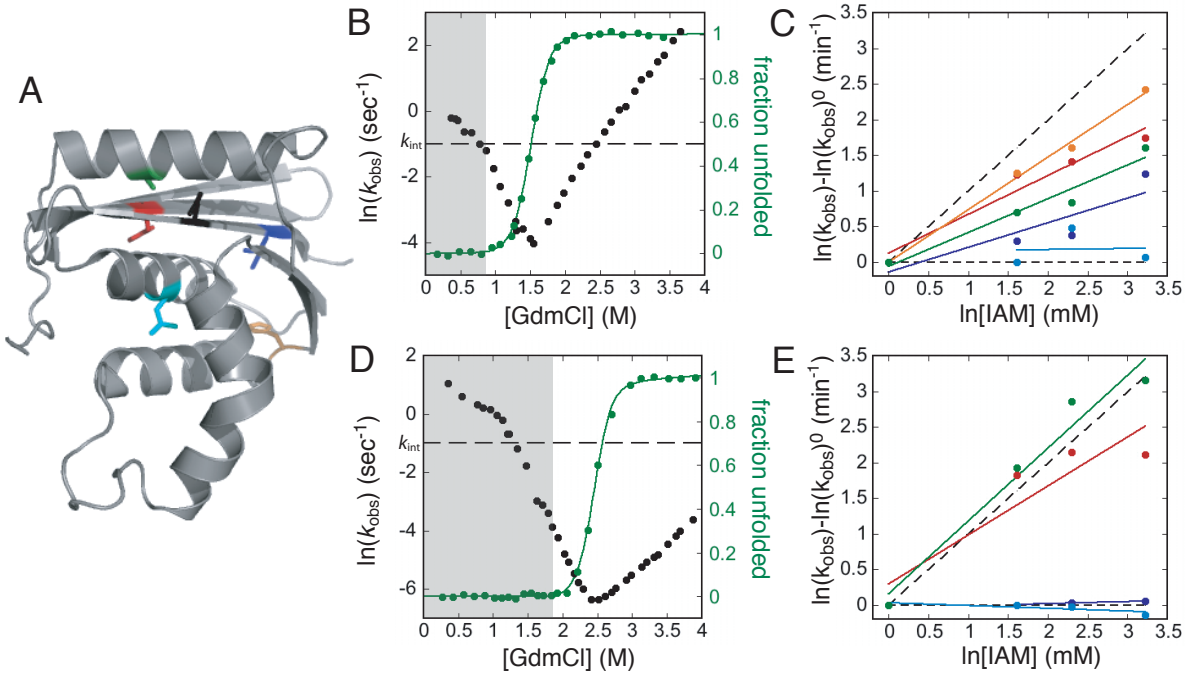
	$\Delta G°_{unf}$ (kcal/mol)	*m*-value (kcal/mol M^{-1})
RNase H*	6.8 ± 0.4	4.5 ± 0.3
RNH	6.8 ± 0.5	4.5 ± 0.4
RNH I7C	3.6 ± 0.3	3.0 ± 0.2
RNH T9C	6.2 ± 0.2	4.1 ± 0.1
RNH I53C	6.5 ± 0.4	4.9 ± 0.4
RNH L59C	4.6 ± 0.3	3.7 ± 0.3
RNH A133C	6.5 ± 0.3	4.3 ± 0.1
RNH (pH 5.5)	11.1 ± 0.6	5.0 ± 0.3

4.3.2 Determination of the kinetic exchange regime for RNH

Exchange experiments are generally interpreted based on either the EX1 or EX2 regime. Most previous NSHX studies, including those of RNase H*, were conducted under EX2 conditions ($k_{cl}>>k_{int}$). Thiol exchange experiments, on the other hand, can have much greater flexibility regarding kinetic exchange regimes by altering the concentration and identity of the thiol modifying reagent. Thus, it was necessary to determine the kinetic exchange regime for the proposed NSSX experiments with the alkylating agent iodoacetamide (IAM).

A comparison of the global folding rate as obtained by CD-monitored kinetics, which can be used as a stand-in for the slowest closing rate in the exchange formalism, to the intrinsic alkylation rate for 10 mM IAM reveals that the folding rates at [GdmCl] relevant for NSSX (up to 0.8M) are generally slightly faster than the intrinsic alkylation rate of modification (16) (Figure 1B). This comparison suggests that the system may be biased toward EX2 exchange but is also relatively close to the EXX exchange regime.

Figure 1. (A)*E. coli* ribonuclease H. Positions of cysteine probes used for kinetic regime testing shown in sticks: I7 (blue), Y22 (red), L49 (cyan), H114 (orange), and A137 (green). The D10 side chain is shown in black sticks. (B and D) GdmCl-induced equilibrium denaturation and folding and unfolding kinetics for RNH (B) and D10A (D). The dashed line shows the intrinsic alkylation rate for 10 mM IAM, and the shaded region indicates the [GdmCl] available for NSSX. (C and E)[IAM]-dependence of observed rates for probes in the RNH background at 0.8 M GdmCl(C) and the D10A background at 1.4 M GdmCl (E); colors are the same as in panel A. The black dashed lines with no slope represent expected EX1 behavior; the slope of one represents expected EX2 behavior. In RNH, L49C exchanges by EX1; all others exchange by EXX. In D10A, I7C and L49C exchange by EX1, and Y22C and A137C by EX2.



To confirm the relationship between the folding kinetics as measured by CD and the exchange regime, the dependence of the exchange rate on [IAM] was measured at 0, 0.4, and 0.8 M GdmCl for five cysteine probes (Figure 1A,C; Table 2). Probes with slopes below 0.25 were classified as EX1, slopes above 0.7 EX2, and slopes between 0.5-0.6 EXX. At 0 M GdmCl, four of the probes exchange in EX2. At higher denaturant concentrations, however, the majority of the probes exchange by the EXX mechanism. The predominance of EXX exchange, the difficulty of interpreting such exchange behavior, and the transition between exchange regime over the range of denaturant concentrations at which the experiments would be conducted made this system intractable for further experiments.

Table 2. Results of kinetic exchange regime testing for five cysteine probes in RNH across the range of [GdmCl] available for NSSX.

	0 M GdmCl	0.4 M GdmCl	0.8 M GdmCl
I7C	slow	EXX	EXX
Y22C	EX2	EXX	EXX
L49C	EX2	EX1	EX1
H114C	EX2	EXX	EXX
A137C	EX2	EXX	EXX

4.3.3 Characterization of a stabilized RNH and cysteine variants

Due to the complications of working with RNH for these NSSX studies – population of the folding intermediate upon some cysteine mutations and EXX exchange – a new background was pursued. To accommodate the perturbing effect of cysteine, we chose to work in the background of a previously characterized mutation, D10A, which removes a destabilizing electrostatic interaction from the periphery of the protein, increasing the overall stability while leaving the stability of the intermediate unchanged (21). Furthermore, NSHX experiments have been conducted with this protein (29), making it an excellent candidate for thiol exchange.

A version of *E. coli* RNase H* D10A with an N-terminal 6-His tag and a C-terminal tag for fluorescent labeling was made for thiol exchange studies; this variant is referred to here simply as D10A. The stability of this construct was measured, as well as the stabilities of a number of single cysteine mutants (Table 3). D10A is stabilized by almost 5 kcal/mol relative to RNH, and none of the mutations in the D10A background resulted in a significantly depressed *m*-value, although they do show a range of destabilization, from 0.2 to 4.1 kcal/mol. Therefore, it appears that the increased stability of the D10A parent protein makes it less susceptible to gross disruptions of structure and/or folding behavior due to the introduction of cysteine mutations.

Table 3. Stabilities of D10A *E. coli* RNase H under NSSX conditions with the peptide tags and upon cysteine substitutions at various positions.

	$\Delta G°_{unf}$ (kcal/mol)	*m*-value (kcal/mol M^{-1})	$\Delta\Delta G°$ (kcal/mol)
D10A	11.5 ± 0.6	5.0 ± 0.3	
D10A I7C	9.5 ± 0.4	4.8 ± 0.3	2.0
D10A G23C	8.1 ± 0.9	4.5 ± 0.5	3.4
D10A A24C	11.3 ± 0.5	5.0 ± 0.2	0.2
D10A I25C	7.6 ± 0.3	4.6 ± 0.2	3.9
D10A M47C	7.4 ± 0.4	4.6 ± 0.3	4.1
D10A I53C	10.7 ± 0.4	5.1 ± 0.2	0.8
D10A A55C	10.1 ± 0.5	5.2 ± 0.3	1.4
D10A V65C	9.5 ± 0.5	5.0 ± 0.3	2.0
D10A V74C	9.1 ± 0.8	5.1 ± 0.4	2.4
D10A I78C	8.6 ± 0.4	4.8 ± 0.2	2.9
D10A L107C	8.0 ± 0.5	5.1 ± 0.3	3.5
D10A A137C	8.7 ± 0.3	4.7 ± 0.2	2.8

The folding and unfolding kinetics for this parent protein were also monitored by CD (Figure 1D), and the resulting chevron plot showed the same characteristic shape. The D10A variant slightly speeds folding, but its main effect is slowed unfolding. Importantly, the global closing rates in the folded baseline broadly span the intrinsic modification rate, suggesting that perhaps both the EX1 and EX2 regimes could be accessed.

4.3.4 Determination of kinetic exchange regimes

The kinetic exchange regime was determined for the same five positions evaluated in RNH at 1.4 M GdmCl; for probes that exchanged too slowly to obtain reliable alkylation rates at this denaturant concentration, measurements were repeated at 1.8 M GdmCl (Figure 1E; Table 4). Four of the five probes showed either distinct EX1 or EX2 behavior in the D10A background that was consistent over the [GdmCl] measured; the final probe exchanged too quickly to measure. Thus, introducing the D10A mutation alters the conformational dynamics such that most probes that exchanged in EXX in RNH showed either EX1 or EX2 kinetics in the D10A background. Therefore, all further NSSX experiments were conducted in the background of the D10A mutation (see chapter 5).

Table 4. Results of kinetic exchange regime testing for five cysteine probes in D10A.

	1.4 M GdmCl	**1.8 M GdmCl**
I7C	EX1	EX1
Y22C	EX2	n.d.
L49C	EX1	n.d.
H114C	fast	fast
A137C	EX2	EX2

4.4 Discussion

These preliminary experiments revealed that the original goal of characterizing *E. coli* RNase H by EX2 NSSX would not be possible with the current system. NSSX under EXX exchange conditions could produce interesting results; analysis, however, would be complicated. Alternatively, EX1 or EX2 exchange kinetics could be achieved by altering conditions such as pH or the identity or concentration of the modifying reagent. Even so, the problem remains that some of the mutations appear to result in catastrophic destabilization of the protein, which would interfere with interpretation of NSSX results.

It was shown, however, that increasing the stability of the parent protein – for example, by introducing the stabilizing D10A mutation – can ameliorate this affect. RNH is also significantly more stable at pH 5.5 than at pH 8.6 (Table 1), and decreasing the pH would slow k_{int} for IAM, suggesting that perhaps an EX2 experiment could be conducted with RNH at a decreased pH. It is possible, however, that decreasing the pH could slow alkylation sufficiently to make the experiments practically intractable, since protein incubated too long with IAM may be subject to side reactions that could affect the protein stability and bias the results.

Instead of changing the pH, the protein can be stabilized by introducing the D10A mutation. This mutation also changes the kinetics such that most probes exchange in interpretable regimes (either EX1 or EX2, rather than EXX). This result suggested the D10A variant may yield

interesting results through a mixed EX1/EX2 NSSX experiment; the resulting study is described in detail in the next chapter.

4.5 References

1. Bai Y, Sosnick TR, Mayne L, & Englander SW (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* 269(5221):192-197.
2. Bollen YJ, Kamphuis MB, & van Mierlo CP (2006) The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* 103(11):4095-4100.
3. Chamberlain AK, Handel TM, & Marqusee S (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* 3(9):782-787.
4. Fuentes EJ & Wand AJ (1998) Local dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* 37(11):3687-3698.
5. Krishna MM, Lin Y, Rumbley JN, & Englander SW (2003) Cooperative omega loops in cytochrome c: role in folding and function. *J Mol Biol* 331(1):29-36.
6. Maity H, Maity M, Krishna MM, Mayne L, & Englander SW (2005) Protein folding: the stepwise assembly of foldon units. *Proc Natl Acad Sci U S A* 102(13):4741-4746.
7. Vadrevu R, Wu Y, & Matthews CR (2008) NMR analysis of partially folded states and persistent structure in the alpha subunit of tryptophan synthase: implications for the equilibrium folding mechanism of a 29-kDa TIM barrel protein. *J Mol Biol* 377(1):294-306.
8. Yan S, Kennedy SD, & Koide S (2002) Thermodynamic and kinetic exploration of the energy landscape of Borrelia burgdorferi OspA by native-state hydrogen exchange. *J Mol Biol* 323(2):363-375.
9. Cliff MJ, *et al*. (2006) A thiol labelling competition experiment as a probe for sidechain packing in the kinetic folding intermediate of N-PGK. *J Mol Biol* 364(4):810-823.
10. Feng Z, Butler MC, Alam SL, & Loh SN (2001) On the nature of conformational openings: native and unfolded-state hydrogen and thiol-disulfide exchange studies of ferric aquomyoglobin. *J Mol Biol* 314(1):153-166.
11. Ha JH & Loh SN (1998) Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat Struct Biol* 5(8):730-737.
12. Jha SK & Udgaonkar JB (2007) Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* 282(52):37479-37491.
13. Stratton MM, Cutler TA, Ha JH, & Loh SN (2010) Probing local structural fluctuations in myoglobin by size-dependent thiol-disulfide exchange. *Protein Sci* 19(8):1587-1594.
14. Silverman JA & Harbury PB (2002) The equilibrium unfolding pathway of a (beta/alpha)8 barrel. *J Mol Biol* 324(5):1031-1040.
15. Krishnan B & Gierasch LM (2011) Dynamic local unfolding in the serpin alpha-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nat Struct Mol Biol* 18(2):222-226.
16. Burguete AS, Harbury PB, & Pfeffer SR (2004) In vitro selection and prediction of TIP47 protein-interaction interfaces. *Nat Methods* 1(1):55-60.
17. Isom DG, Vardy E, Oas TG, & Hellinga HW (2010) Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity. *Proc Natl Acad Sci U S A* 107(11):4908-4913.

18. Johnson CP, Tang HY, Carag C, Speicher DW, & Discher DE (2007) Forced unfolding of proteins within cells. *Science* 317(5838):663-666.
19. Silverman JA & Harbury PB (2002) Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J Biol Chem* 277(34):30968-30975.
20. Dabora JM & Marqusee S (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: characterization of a partially folded state. *Protein Sci* 3(9):1401-1408.
21. Raschke TM, Kho J, & Marqusee S (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* 6(9):825-831.
22. Raschke TM & 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(4):298-304.
23. Connell KB, Miller EJ, & 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(2):450-460.
24. Goedken ER & Marqusee S (2001) Co-crystal of Escherichia coli RNase HI with Mn2+ ions reveals two divalent metals bound in the active site. *J Biol Chem* 276(10):7266-7271.
25. Connell KB, Horner GA, & 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(2):461-470.
26. Dabora JM, Pelton JG, & Marqusee S (1996) Structure of the acid state of Escherichia coli ribonuclease HI. *Biochemistry* 35(37):11951-11958.
27. Park C & Marqusee S (2004) Probing the high energy states in proteins by proteolysis. *J Mol Biol* 343(5):1467-1476.
28. Cecconi C, Shank EA, Bustamante C, & Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. *Science* 309(5743):2057-2060.
29. Goedken ER & Marqusee S (2001) Native-state energetics of a thermostabilized variant of ribonuclease HI. *J Mol Biol* 314(4):863-871.
30. Berger A & Linderstrom-Lang K (1957) Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* 69:106-118.
31. Hvidt A & Nielsen SO (1966) Hydrogen exchange in proteins. *Adv Protein Chem* 21:287-386.
32. Xiao H, *et al*. (2005) Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment. *Protein Sci* 14(2):543-557.
33. Yin J, *et al*. (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* 102(44):15815-15820.
34. Santoro MM & Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27(21):8063-8068.
35. Spudich G & Marqusee S (2000) A change in the apparent m value reveals a populated intermediate under equilibrium conditions in Escherichia coli ribonuclease HI. *Biochemistry* 39(38):11677-11683.

Chapter 5

Mapping protein side-chain exposure on the energy landscape: mixed EX1/EX2 native-state alkyl-proton exchange

Adapted from Bernstein R, Schmidt KL, Harbury PB & Marqusee S (2011) *PNAS* (under review)

5.1 Introduction

In order to fold and function, proteins must explore structural fluctuations away from their native conformation. Such fluctuations result in a distribution of conformations of differing stabilities that, together with the barriers separating them, constitute the energy landscape. These high-energy partially unfolded forms are important for various biological functions, including allostery (1-3), catalysis (4), motions of motor proteins (5-7), and aggregation (8-10).

In spite of their importance, unfolding events occur very rarely under native conditions, and the populations of these conformations are very small, rendering experimental characterization of such partially unfolded species particularly challenging. Native-state amide hydrogen exchange (NSHX) has proven to be a powerful technique to identify and provide residue-specific structural information about such species (11). The power of NSHX to interrogate rare conformations is based on the Linderstrom-Lang model for the exchange process (12), where an amide hydrogen in the "closed" or native conformation must undergo a fluctuation to some alternative "open" conformation in order to exchange:

$$closed \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} open \xrightarrow{k_{int}} exchanged \quad \text{(scheme 1)}$$

Because the majority of molecules are in the native conformation, they are not available for exchange, allowing for the detection of rare, high-energy "open" species.

Depending on the relative kinetics of these processes, the observed exchange rate (k_{obs}) can report on either the rate of forming the open state, k_{op} (EX1 regime, where $k_{cl} << k_{int}$), or on the stabilities of open forms, K_{op} (EX2 regime, where $k_{cl} >> k_{int}$) (13). Under native conditions and at neutral pH, HX generally proceeds by the EX2 mechanism. Therefore, NSHX, which has successfully detected high-energy partially unfolded conformations in a number of proteins (see, for example, (14-21)), usually provides information only about the equilibrium populations of these species and not the barriers between them. Recently, methods employing triplet-triplet energy transfer have been used to probe both the kinetics and thermodynamics of extremely fast fluctuations on the native side of the rate-limiting barrier (22). Other studies have used a combination of EX1 and EX2 HX data to localize intermediates along the reaction coordinate (15, 21, 23-27), but such experiments require probing the system under different experimental conditions, such as different pHs, and that these changes do not alter the protein's behavior.

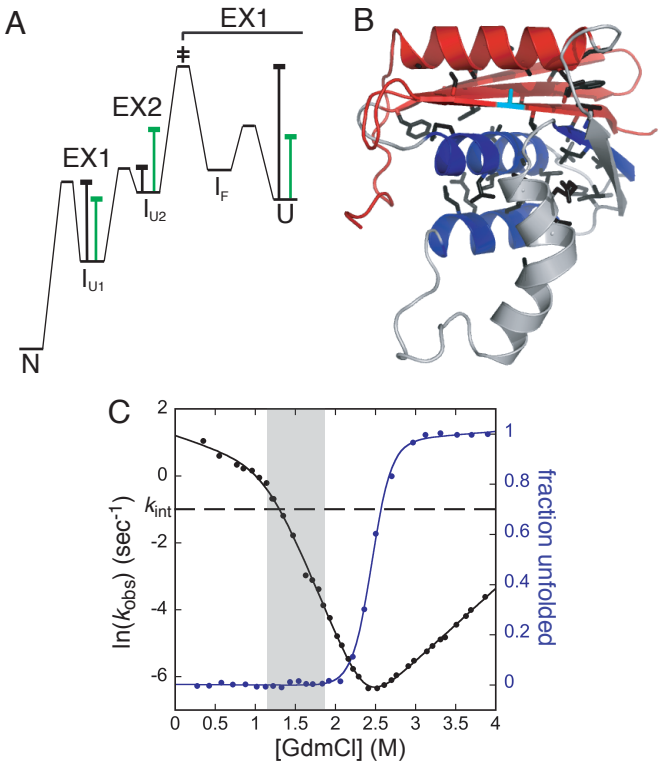
Thiol alkyl-proton exchange (SX), which monitors cysteine side-chain modification, can access both the EX1 and EX2 regimes by simple adjustment of the concentration and chemical identity of the thiol-modifying reagent. Furthermore, under appropriately chosen experimental conditions, cysteine probes may kinetically partition into different exchange regimes, which allows localization of the associated exchange conformations on the reaction coordinate. For example, as shown in Figure 1A, if the intrinsic thiol modification rate is set such that global unfolding probes exhibit EX1 exchange kinetics, then so must all probes modified through a state on the unfolded side of the rate-limiting barrier. Moreover, the modification rates of such probes must match the rate of global unfolding. This tenet holds because all probe modification events are at least as fast as modification through the unfolded state, which itself is faster than re-crossing the rate-limiting barrier from the unfolded side back to the native side. A corollary is

that probes exhibiting EX2 exchange kinetics must be modified through states located on the native side of the rate-limiting barrier. It is also possible for probes modified on the native side to exhibit EX1 exchange kinetics, but such cases can be distinguished from unfolded-side probes because their rates of exchange will exceed the global unfolding rate. Thus the kinetic regime (EX1 or EX2) and exchange rate of each cysteine probe report on the position of its exchange state relative to the dominant feature in the energy landscape.

In addition to the flexibility of SX experiments to access kinetic exchange regimes denied to HX, SX also has the power to probe important side-chain packing interactions that can be invisible to HX experiments. This ability depends on the presence of cysteine residues, which presents a potential drawback, as such experiments are either limited to naturally occurring cysteines or require the introduction of non-native cysteines that may result in destabilization or changes to protein behavior. Thiol exchange, however, has been used to monitor the energetics of conformational changes such as unfolding (28-36), suggesting that the perturbing effect of cysteine substitutions can be overcome.

Here we carry out a detailed investigation of native-state thiol alkyl-proton exchange (NSSX) for a large set of cysteine residues throughout a variant of *E. coli* ribonuclease H (RNase H; Figure 1B). This 155-amino acid protein is a well-characterized model system known to populate a kinetic folding intermediate that resembles a high-energy conformation detected by NSHX (16, 37, 38). To expedite the SX characterization, we employ a modified version of Silverman's gel-based methodology to simultaneously monitor alkylation at multiple sites (39). The folding kinetics of the previously characterized variant D10A (40) relative to the intrinsic alkylation rate with the modifying agent iodoacetamide (IAM) allow us to apply a mixed EX1/EX2 approach to identify species on both sides of the rate-limiting barrier in a single experiment, characterize novel partially unfolded forms present on the folded side, and interrogate the rate-limiting folding barrier at the residue level.

Figure 1. Kinetic partitioning and D10A *E. coli* RNase H. (A) A model energy landscape showing the kinetic partitioning effect. Barrier heights are hypothetical and represented by vertical lines (black, conformational closing; green, intrinsic chemical modification). In this simple landscape, species on the unfolded side of the rate-limiting barrier exchange in EX1 and report on the global unfolding rate. Probes modified through fluctuations on the native side of the barrier may exchange in either EX1 or EX2. EX1 probes on the native side will be modified with rates faster than global unfolding. EX1 probes on each side of the transition state can therefore be distinguished by their observed kinetic behavior. (B) D10A *E. coli* RNase H. Positions investigated in NSSX experiments are shown as black sticks; Ala10 is shown in cyan sticks, and the backbone is colored according to the model of the NSHX PUF, with the structured region in blue and the unstructured in red. (C) GdmCl-induced equilibrium denaturation and unfolding kinetics for D10A. The dashed line shows the intrinsic alkylation rate for 10 mM IAM, and the shaded region indicates [GdmCl] used for NSSX with EX1 probes.



5.2 Methods

5.4.1 Gene construction and protein production

An N-terminal 6-His tag and a C-terminal ybbR tag (42) were added to the gene encoding the D10A *E. coli* RNase H*variant (40). Individual cysteine mutations were introduced by QuikChange mutagenesis.

All proteins were expressed in BL21(DE3) pLysS *E. coli* cells grown in LB medium. For alkylation experiments, single colonies of three to eight single-cysteine variants were grown separately overnight and LB medium was inoculated with culture from each of these variants.

Expression was induced at OD 0.6-0.8 with 1 mM IPTG and cells were harvested after 2-3 hours.

To obtain pure protein for circular dichroism (CD) experiments, cells were resuspended in 50 mM Tris pH 8, 50 mM NaCl, sonicated, and run over a heparin column with a gradient to 500 mM NaCl. Fractions containing the protein were pooled and subsequently loaded on a nickel column, washed with 50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole, and eluted at 500 mM imidazole. For single-cysteine mutants, all buffers also contained 1 mM TCEP. Pure protein was then dialyzed into CD buffer (20 mM Tris pH 8.6, 50 mM KCl; for single-cysteine mutants, buffer also contained 0.1 mM TCEP). To purify protein for alkylation experiments, the heparin column was omitted. Cells were resuspended in 50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole, 1mM TCEP, sonicated, and purified on a nickel column. Protein was then dialyzed into alkylation conditions (100 mM bicine pH 8.6, 50 mM KCl, 1 mM TCEP).

5.2.2 Thiol alkyl-proton exchange

Pools of single-site cysteine mutants were grown and purified in eight bins. Thiol exchange experiments were conducted in 100 mM bicine pH 8.6, 50 mM KCl, 1 mM TCEP (CD experiments were conducted in 20 mM Tris pH 8.6, 50 mM KCl, and 0.1 mM TCEP for single-cysteine mutants). Samples were prepared by a modified scheme based on that originally used by Silverman and Harbury (39). Protein was labeled with fluorescein-CoA by the enzyme Sfp phosphopantetheinyl transferase, denatured in 4 M GdmCl pH 8.6, cyanylated with 25 mM dinitrothiocyanobenzoic acid (NTCB), and precipitated with 0.5% trichloroacetic acid. Pellets were washed twice with acetone, resuspended in 8 M urea in 1 M ammonium hydroxide, and incubated at room temperature for 1 hr, resulting in cleavage at the positions of cyanylation. Fragments were separated by gel electrophoresis and fluorescein-CoA-labeled protein was visualized with a Typhoon fluorescence imager. Band intensities were quantified using ImageQuant and normalized based on a syntaxin loading control. Normalized data were fit to a three-parameter single exponential using Kaleidagraph.

To determine the kinetic exchange regime, pools of cysteine-containing D10A variants were incubated with 1, 5, 10, and 25 mM iodoacetamide (IAM) at 1.4 and 1.8M GdmCl; alkylation was quenched at various time points with an excess of DTT. Probes with slopes below 0.25 were classified as EX1, slopes above 0.7 EX2, and slopes between 0.5-0.6 EXX. Subsequent NSSX experiments were conducted in an analogous method by measuring alkylation rates as a function of [GdmCl] using 10 mM IAM.

Thiol alkylation studies were analyzed using the classic EX1/EX2 formalism applied to hydrogen exchange. Data from EX2 probes were fit with a model that allows for both denaturant-dependent and denaturant-independent:

$\Delta G = - RT \ln (\exp((m[GdmCl] - \Delta G°_{SX})/RT) + \exp(-\Delta G°_{fluc}/RT))$

In all but one case, the data were fit sufficiently well with just one of the terms, and this equation could be simplified to either of the two following equations:

$\Delta G = - RT \ln (\exp((m[GdmCl] - \Delta G°_{SX})/RT))$

$\Delta G = - RT \ln (\exp (-\Delta G°_{fluc}/RT))$

5.2.3 Circular dichroism

For equilibrium melts, CD data were collected using an Aviv 410 spectropolarimeter with a Peltier temperature-controlled sample holder and an Aviv 62A DS spectropolarimeter with a HP temperature-controlled sample holder, both with a 1 cm path length cuvette. Protein denaturation studies were conducted by monitoring the ellipticity at 222 nm as a function of [GdmCl] at 25ºC. Samples were allowed to equilibrate overnight, and the signal for each sample was averaged over 60 seconds. Free energies of unfolding and *m*-values were calculated using Kaleidagraph, assuming a two-state transition and linear dependence on denaturant(48). For chevron plots, an Aviv 202 spectropolarimeter with a stopped-flow system and an Aviv 410 spectropolarimeter were used. Ellipticity at 222 nm was measured as a function of time. Experiments done on the 202 were repeated seven times and averaged. Data collected on the 410 were taken once and spot-checked for reproducibility.

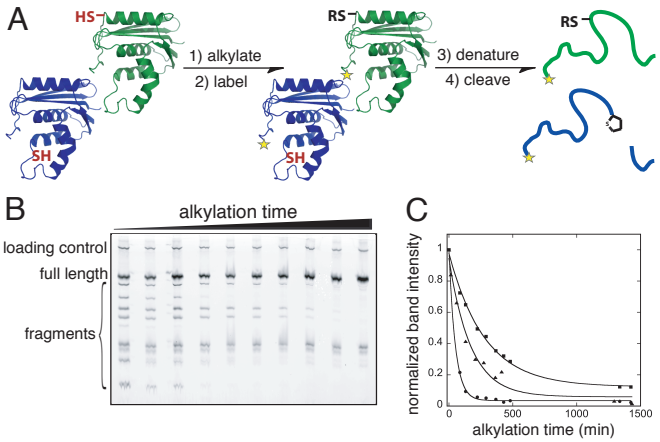
5.2 Results and Discussion

5.2.1 Tuning kinetics to achieve EX1/EX2 kinetic partitioning

A version of the well-characterized variant D10A *E. coli* RNase H* (* denotes a variant of the protein with the three native cysteines mutated to alanine (41)) suitable for our NSSX experiments was created by the addition of an N-terminal 6-His tag and a C-terminal ybbR tag for fluorescent labeling(42); this variant is referred to here simply as D10A. These tags do not affect the stability of RNase H* as measured by equilibrium GdmCl denaturation (see Chapter 3, Table 1). To explore the relative kinetics of D10A folding and alkylation, the folding and unfolding kinetics were determined by CD. The resulting chevron plot reveals the characteristic rollover in the folding limb attributed to the presence of a folding intermediate, as expected based on previous folding studies of *E. coli* RNase H* variants carried out under slightly different conditions(40). The folding kinetics show that, under the conditions of the thiol exchange experiments for the slowest-exchanging probes (~1.2 – 1.8 M GdmCl), k_{int} is slightly faster than k_f (Figure 1C).

To determine the kinetic exchange regime for each probe, the dependence of the observed modification rates on the concentration of modifying reagent was measured; in EX2, the dependence is linear, while EX1 exchange is independent of modifying reagent concentration. Cysteine modification was detected by a scheme based on that originally used by Silverman (39)(Figure 2). Briefly, pools of single-cysteine variants that were exposed to iodoacetamide for various lengths of time were subsequently fluorescently labeled, denatured, cyanylated at unalkylated cysteines, and cleaved at the positions of cyanylation. Fragments were separated by gel electrophoresis and normalized fluorescent band intensities were fit to a three-parameter single exponential decay to yield an observed exchange rate.

Figure 2. A schematic diagram of the NSSX experiment. (A) Single-cysteine variants are pooled and exposed to a thiol-reactive modifying agent; buried thiols (blue protein) are alkylated less quickly than exposed thiols (green protein). Proteins are then fluorescently labeled at the C-terminal peptide tag, unfolded, and exposed to the cyanylating agent NTCB. Previously unreacted cysteines are cyanylated and cleaved upon pH increase, while modified cysteines do not react with NTCB and thus are not cleaved. (B) Peptide fragments are resolved using a high-resolution polyacrylamide gel and band intensities are normalized based on a loading control and fit to a single exponential. (C) Representative traces for the Y22C probe at multiple [GdmCl] (squares, 0.71 M; triangles, 1.09 M; circles, 1.30 M); alkylation rates increase with increasing [GdmCl].



Using this protocol, the kinetic exchange regime was characterized for 43 cysteine probes introduced by site-directed mutagenesis into the D10A background (Figure 3A, B, Table 1). Seven probes exchanged in the EX1 regime and 19 in the EX2 regime, demonstrating that successful kinetic tuning had been achieved. Five probes exchanged by a mixture of EX1 and EX2 termed EXX(43), indicating that the closing rate for these probes is approximately the same as the intrinsic alkylation rate, and 12 exchanged too quickly to measure.

Figure 3. NSSX data and results. (A) The kinetic exchange regime was determined for all probes; a representative sample at 1.4 M GdmCl is shown. The dashed line with no slope represents expected EX1 behavior and the dashed line with unity slope shows expected EX2 behavior. (B) D10A with NSSX probes colored according to their kinetic exchange regime (EX1 blue, EX2 red, EXX purple). Positions that exchange fast or with fluctuations are not shown. (C) Alkylation rates for EX1 probes plotted as a function of [GdmCl] relative to the fit of the CD chevron; the dashed line is the extrapolation of the chevron unfolding limb. The error for the fit of the exponential decay for each point is at most ±0.5 ln(k_{obs}), with an average error of ±0.2 ln(k_{obs}). (D) EX1 probes, shown in blue spheres, are mostly in the protein core. (E) Representative samples of denaturant-dependent EX2 probes, colored by PUFs as designated in Table 2. The green dashed line reflects a linear extrapolation of the denaturant-dependent exchange region for a probe with denaturant-independent exchange at low [GdmCl]. The black dashed line is an extrapolation of the kinetic ΔG and *m*-value as measured by CD. The error for the fit of the exponential decay for each point is at most ±0.3 kcal/mol, with an average error of ±0.15 kcal/mol. (F) PUFs as represented by SX probes rendered in spheres: green, helix E PUF; yellow, clamshell PUF; cyan, high-energy PUF. EX1 probes are shown as blue sticks.

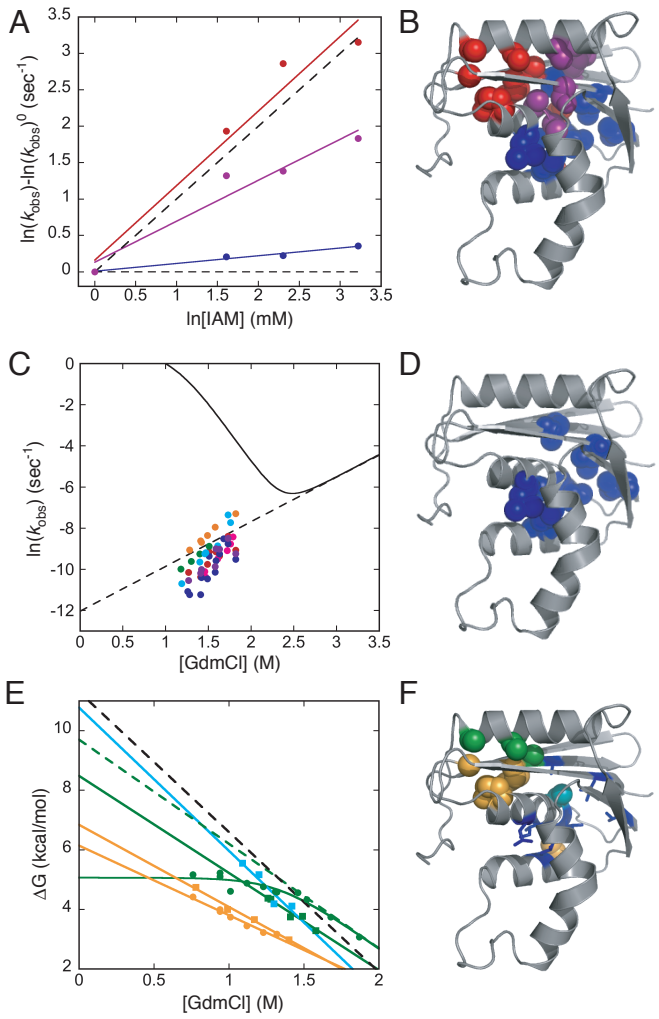


Table 1. NSSX results for all probes. Errors have been propagated including the average error of each point and the overall error of the fit. Values for ΔG_{fluc} were determined assuming *m*-values of zero.

	Secondary structure	Behavior	$\ln(k_{obs})(H_20)$	$\Delta G°_{SX}$ (kcal/mol)	*m*-value (kcal/mol M^{-1})	$\Delta G°_{fluc}$ (kcal/mol)
I7	strand 1	EX1	-14.3 ± 0.6		1.8 ± 0.4	
F8	strand 1	fast				
T9	strand 1	EXX				
G11	strand 1	fast				3.4 ± 0.5
S12	strand 1	EX2				3.5 ± 0.2
G20	strand 2	EX2		6.2 ± 0.4	2.3 ± 0.3	
G21	strand 2	fast				
Y22	strand 2	EX2		8.1 ± 0.6	3.0 ± 0.5	
G23	strand 2	EX2		10.6 ± 0.6	3.6 ± 0.4	
A24	strand 2	EX1	-18.2 ± 1.0		3.2 ± 0.6	
I25	strand 2	EXX				
S36	strand 3	fast				
A37	strand 3	EX2				3.1 ± 0.3
Y39	strand 3	fast				
M47	helix A	EX2		6.7 ± 0.3	2.7 ± 0.2	
L49	helix A	EX1	-14.0 ± 1.1		2.0 ± 0.8	
A51	helix A	EXX				
A52	helix A	EXX				
I53	helix A	EX1	-17.0 ± 0.8		2.7 ± 0.4	
V54	helix A	EX2				2.9 ± 0.3
A55	helix A	EX2		10.3 ± 0.7	4.4 ± 0.5	
L56	helix A	EX1	-17.4 ± 0.6		3.3 ± 0.4	
L59	helix A	EX2				3.0 ± 0.2
V65	strand 4	EX1	-14.7 ± 0.7		2.0 ± 0.4	
L67	strand 4	EX2				3.3 ± 0.1
S68	strand 4	EX2				4.3 ± 0.3
V74	helix B	EX2				4.4 ± 0.4
R75	helix B	EX2				2.6 ± 0.2
I78	helix B	EX2				3.6 ± 0.3
I82	helix C	fast				
L103	helix D	EX2				2.3 ± 0.1
R106	helix D	fast				
L107	helix D	EX1	-13.4 ± 0.7		2.0 ± 0.5	
A110	helix D	EX2		8.8 ± 0.6	2.4 ± 0.4	
H114	loop	fast				
I116	strand 5	EX2				4.9 ± 0.4
V121	strand 5	fast				
E129	helix E	fast				
N130	helix E	fast				
A133	helix E	EXX				
L136	helix E	fast				
A137	helix E	EX2		8.8 ± 0.4	3.5 ± 0.3	
A141	helix E	EX2		9.7 ± 0.8	3.5 ± 0.5	5.1 ± 0.1

The perturbing effect of introducing cysteine mutations was investigated by determining the $\Delta G°_{unf}$ with CD-monitored GdmCl-induced denaturation for 12 single-cysteine variants in the D10A background. Destabilization due to the cysteine mutations varied ($\Delta\Delta G°$= 0.2 – 4.2 kcal/mol) (see Chapter 4, Table 3), indicating that caution must be used when interpreting the site-specific energetics determined by SX.

Having identified suitable conditions where probes partition between EX1 and EX2 kinetics and defined a range of destabilization due to the cysteine mutations, NSSX (thiol reactivity as a function of [GdmCl]) was carried out on D10A. The following analysis focuses only on probes that reacted by either EX1 or EX2 kinetics; due to the complexity of interpreting the exchange behavior for EXX probes, they were omitted from further structural interpretation.

5.3.2 EX1 probes reveal information about early events in folding

Under the conditions of our experiment, we expect that all modification on the unfolded side of the barrier, either from an intermediate or from the unfolded state, should exhibit EX1 kinetics. The rate of modification for unfolded-side probes should then correspond to the global unfolding rate. As expected, the EX1 alkylation rates observed under the conditions of our experiment agree with the expected global unfolding rate as determined by CD (Figure 3C). A linear extrapolation to 0 M GdmCl, however, yields a range of $k_{op}(H_2O)$ from about 10^{-6} sec^{-1} to 10^{-8} sec^{-1}, mostly due to variation in the calculated slopes, or *m*-values (Table 1). Similar variation in the opening rates determined by HX has been seen for the protein turkey ovomucoid third domain and interpreted as independent unfolding events (23). It is unclear, however, whether the distribution seen in these NSSX experiments represents a real distribution in residue-specific opening rates, a result of small changes due to the effect of the cysteine substitution, or simply noise in the data.

Most of the probes that show EX1 behavior reside in the core of the protein (Figure 3D), identified both as an early protected region (<10 msec) in pulse-labeling HX studies (38) and an equilibrium intermediate by NSHX (16, 37). The exchange behavior of these probes indicates that, as expected, this core region is not modified upon fluctuations on the native side of the barrier, and the protein must completely unfold for chemical modification of these sites at both the side-chain and backbone level.

Two EX1 probes, I7C and A24C, fall outside the core as previously defined by HX experiments. Our model, based on both pulse-labeling and native-state hydrogen exchange data, suggests that these residues are exposed in the folding intermediate. There were no data, however, to indicate where on the folding trajectory these residues become protected. Their NSSX EX1 behavior and observed alkylation rates, which match those for the other EX1 probes, suggest that these two side chains only become exposed and available for modification on the unfolded side of the barrier. For further confirmation, we measured the effects of the cysteine substitution on the folding kinetics. While A24C has no effect, the I7C mutation slows folding (Figure 4), and analysis of the kinetic effect indicates that the side chain gains structure on the unfolded side of the barrier, primarily after formation of the intermediate but before the transition state (Φ^{I}=0.3, $\Phi^{\ddagger}$=1.2 (Table 2)). Thus, modification of I7C likely occurs on the unfolded side of the barrier. The side chains of I7 and A24 are directed toward the core region, consistent with the suggestion that both are involved in important, previously unidentified packing interactions that stabilize the transition state.

Figure 4. Chevron plots for I7C(D10A) (blue circles), A24C(D10A) (blue squares), and A55C(D10A) (red circles) relative to D10A fit (black line).

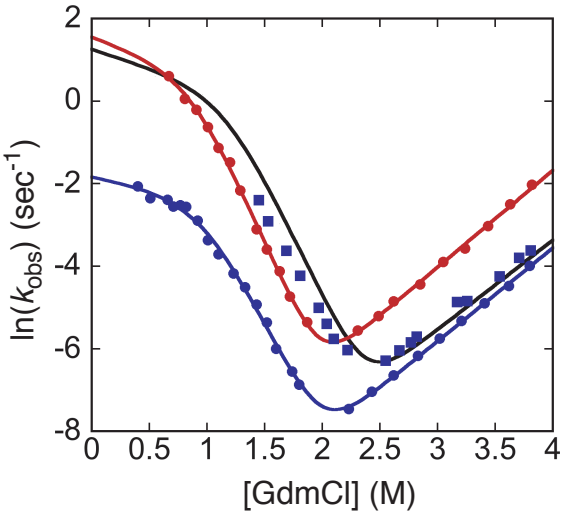


Table 2. Results of kinetic fits for D10A and cysteine variants fit to a three-state, on-pathway intermediate model.

	D10A	**D10A I7C**	**D10A A55C**
$\Delta G°_{GdmCl\ melt}$ (kcal/mol)	11.5 ± 0.6	9.5 ± 0.4	11.3 ± 0.5
$m_{GdmCl\ melt}$ (kcal/mol M^{-1})	5.0 ± 0.3	4.8 ± 0.2	5.0 ± 0.2
$\Delta G°_{kinetic}$ (kcal/mol)	11.3	9.2	10.2
$m_{kinetic}$ (kcal/mol M^{-1})	4.7	4.5	5.1
$\Delta G°_{ui}$ (kcal/mol)	3.4	2.8	2.7
K_{ui}	304	109	95
m_{ui} (kcal/mol M^{-1})	2.9	2.9	3.1
$k_{in}(H_20)$ (sec^{-1})	3.50	0.16	4.76
m_{in} (kcal/mol M^{-1})	0.53	0.33	0.61
$k_{ni}(H_20)$ (sec^{-1})	5.8 x 10^{-6}	3.4 x 10^{-6}	1.6 x 10^{-5}
m_{ni} (kcal/mol M^{-1})	1.3	1.3	1.4
Φ^{I}		0.3	0.6
$\Phi^{‡}$		1.2	0.5

5.3.3 EX2 probes explore ruggedness of the native side of the barrier

For those probes that were modified by EX2 kinetics, the observed alkylation rates were converted to free energies ($\Delta G°_{SX}$) and plotted as a function of [GdmCl] (Figure 3E). As seen in similar native-state hydrogen exchange experiments, these data revealed both denaturant-dependent and denaturant-independent behavior (16, 34, 37). All but one of the SX probes displayed only one of these mechanisms and thus could be fit linearly (A141C required a more complex fit to accommodate both types of alkylation). The resulting *m*-values and extrapolated $\Delta G°_{SX}(H_2O)$ and $\Delta G°_{fluc}$ are shown in Table 1. Eleven of the EX2 probes, all of which reside near the surface of the protein, show alkylation rates with no or low denaturant dependence (*m*-value < 1.4); modification at these positions was thus attributed to local structural fluctuations. Exchange of one additional probe, G11C, was also attributed to local structural fluctuations, but

reacted too quickly under the conditions used for kinetic regime testing to confirm whether its exchange occurred by EX2. The remaining eight EX2 probes had denaturant-dependent alkylation rates, suggesting unfolding events that expose specific regions of the protein, or so-called partially unfolded forms (PUFs). The variation in destabilizing effects of the cysteine mutations prohibits our use of the $\Delta G°_{SX}(H_2O)$ as a criterion to cluster these probes into PUFs; instead, probes were grouped based on their calculated *m*-values and structural relationships, resulting in three clusters (Figure 3F, Table 3).

Table 3. Results of clustering for EX2 probes. Errors for averages are standard deviations.

Cluster	Position	$\Delta G°_{SX}$ (kcal/mol)	*m*-value (kcal/mol M^{-1})
Yellow	G20	6.2 ± 0.3	2.3 ± 0.5
	Y22	8.1 ± 0.6	3.0 ± 0.5
	M47	6.7 ± 0.3	2.7 ± 0.2
	A110	8.8 ± 0.6	2.4 ± 0.4
	Average	7.5 ± 0.6	2.6 ± 0.5
Green	G23	10.6 ± 0.6	3.6 ± 0.4
	A137	8.8 ± 0.4	3.5 ± 0.3
	A141	9.7 ± 0.8	3.5 ± 0.5
	Average	9.7 ± 0.8	3.5 ± 0.5
Cyan	A55	10.3 ± 0.7	4.4 ± 0.5

The PUF with an intermediate *m*-value (Figure 3F, green spheres) consists of side chains that reside on or pack with the final helix of the protein, the E helix. These probes would be exposed in a conformation with the E helix either undocked or unfolded, whose existence has been suggested by several independent lines of evidence. Recent mechanical unfolding studies of *E. coli* RNase H have identified an unfolding intermediate with an unstructured E helix (JWD, manuscript in preparation), and a fragment without the E helix shows cooperative folding and the ability to bind a peptide corresponding to the E helix (44). Furthermore, NSHX on *T. thermophilus* RNase H, a thermophilic homolog of the *E. coli* protein, revealed a PUF corresponding to the unfolding of this helix (45).

These thiol exchange experiments, however, are the first to suggest the presence of this intermediate in full-length *E. coli* RNase H, perhaps because previous *E. coli* RNase H NSHX experiments with both D10A and the wild-type protein were limited to the EX2 regime(16, 37). The hydrogen exchange results reveal the lowest energy PUF involving the E helix, and under the conditions of that experiment, the E helix exchanges from a conformation whose structure correlates with the folding intermediate, with both the E helix and the β sheet unfolded. The conformation suggested by these thiol exchange experiments, with the E helix exposed on the native side of the barrier, would be missed in the NSHX experiments if it were higher in energy than the folding intermediate. Because in our NSSX experiments the well-characterized folding intermediate is accessed more slowly (EX1) than the native-side PUF (EX2), we are able to detect the PUF.

The denaturant dependence, or *m*-value, for this novel unfolding intermediate of 3.5 kcal/mol•M as measured by thiol exchange is quite high relative to the *m*-values of 5.0 kcal/mol•M for global unfolding (equilibrium denaturation monitored by CD), 1.3 kcal/mol•M for unfolding from the native state to the transition state, and 1.8 kcal/mol•M for N to the folding intermediate (CD-monitored kinetics). Such unexpectedly high *m*-values are also seen for the other EX2 probes. If these *m*-values determined by our thiol exchange studies correspond to changes in exposed surface area, the data present a paradox: native-side PUFs that are more solvent-exposed than the transition state for unfolding. This would suggest either that *m*-value is not a good model for progress along the unfolding reaction coordinate, or that we have identified off-pathway intermediates with high degrees of solvent-exposed surface area.

Interestingly, NSHX-monitored unfolding of helix E and the β sheet of RNase H D10A also exhibits an *m*-value of about 3.5(37). One possible explanation for our data is that, although the probes that define this PUF are located along the E helix interface, the structure of this PUF corresponds to unfolding of more than just that helix. In such a scenario, the native-state thiol exchange experiments may access a native-side PUF with the E helix and β sheet unfolded. An alternative explanation is that the free energies derived from these thiol-modification rates increase with denaturant concentration independently of changes in solvent-exposed surface area. If this were the case, our analysis would lead to anomalously high *m*-values (and corresponding ΔG°_{unf}) unrelated to protein conformational change.

A second PUF consists primarily of side chains at the interface of helix A and strand 2 (Figure 3F, yellow spheres). The smaller *m*-values for these probes (m_{avg}= 2.6 kcal/mol•M) suggest less exposure of non-polar surface area in this PUF, perhaps indicating an intermediate formed by a clamshell-like motion in which the β sheet and E helix move up and away from the core helices without unfolding. HX experiments do not reveal any such motion, likely due to the fact that this conformation exposes side chains while retaining the backbone hydrogen-bond network. Furthermore, the side chains exposed in the yellow PUF are proximal to the active site, suggesting that this species may be important for enzymatic activity. Detection of this clamshell-like motion highlights the ability of SX to provide important insights by probing side-chain packing.

The final PUF consists of a single probe, A55C, at the end of helix A (Figure 3F, cyan spheres). This probe resides within the core region as defined by backbone NSHX and is modified with a ΔG°_{SX} and *m*-value close to those associated with global unfolding. This alkylation behavior suggests that, despite its EX2 behavior, this probe may report on global unfolding due to a mutational effect invalidating the kinetic partition. Measurement of the refolding kinetics by CD, however, shows that the mutation does not significantly affect the folding rate (Figure 4, Table 3), which together with the EX2 kinetics implies that alkylation at this position does not require crossing the rate-limiting folding barrier. Thus, we believe that A55C is exposed in a high-energy fluctuation on the folded side of the barrier. It is difficult, however, to construct a structural model for this species based on a single probe. Similarly, A110 C showed analogous behavior to the yellow PUF, but because of its distant location at the end of helix D, this probe was omitted from structural interpretation.

5.3.4 Thiol alkylation provides information about side-chain packing

These SX experiments highlight the importance of side chain-interactions in folding. The EX1 kinetics that report on the global unfolding rate at positions 7 and 24, in conjunction with the Φ-value analysis for I7C, indicates that, during folding, these side chains become structured prior to the rate-limiting transition state, earlier than detected by previous kinetic HX experiments. The clamshell-like (yellow) PUF suggests a novel species, potentially involved in activity, that does not involve any change in backbone hydrogen bonding, and there are likely similar motions in other systems that are important for activity but are largely invisible to HX experiments due to an absence of associated backbone perturbations. Furthermore, analyses of the positions that react by fluctuations (independent of [GdmCl]) or are alkylated too quickly to measure provide additional evidence of variation between side chain and backbone exchange behavior. There are many instances where the side chain is exposed to solvent while the backbone is protected, but there are also four probes (L49C, I116C, A133C, and A137C) that show protection by SX despite exchanging too quickly to measure by HX. These probes do not describe a structurally contiguous region, but their SX behavior indicates that side chains can play a crucial role in anchoring different regions of the protein, even while the backbone is subject to fluctuations.

The nature of intermediates on the native side of the unfolding barrier has been a topic of recent interest. There is some evidence that unfolding begins with the formation of a dry molten globule (22, 46), and Loh and coworkers (47) have shown that SX experiments carried out with thiol-modifying reagents of different sizes can yield information about the magnitude of opening events that expose specific thiol groups. While the intermediates we observe may correspond to such a dry molten globule state, our data only provide information about side-chain exposure for alkylation by iodoacetamide and, by comparison to NSHX, the relative exposure of amide hydrogens. Thus, further experiments are needed to test if any of these newly identified intermediates on the native side of the barrier represent dry molten globules.

5.4 Conclusions

Native-state alkyl-proton exchange offers a powerful complement to hydrogen exchange studies of a protein's energy landscape. In addition to providing information about the environment of the side chain, SX provides the opportunity to obtain both kinetic (EX1) and thermodynamic (EX2) information under the same conditions (pH, temp, etc.). By tuning the intrinsic thiol-alkylation rate and using a well-characterized variant of *E. coli* RNase H, we have assigned probes to alkylation-competent species on either side of the barrier. Furthermore, these SX experiments identify novel PUFs on the folded side of the barrier that in NSHX may have been masked by the lack of backbone exposure or the limitations of an all-EX2 equilibrium experiment.

As a general tool, the approach of a mixed EX1/EX2 thiol exchange experiment has a number of advantages over a traditional EX2 experiment. The ability to assign species to specific regions of the reaction coordinate relative to the rate-limiting barrier is of paramount importance for characterizing folding and unfolding pathways. The kinetic partitioning principle can also be extended by taking advantage of the range of intrinsic rates furnished by a variety of modification agents to explore specific regions of the landscape based on different kinetic barriers. The ability of these NSSX experiments to detect high energy species on the folded side of the rate-limiting barrier, even if the same positions are also exposed in lower energy species

on the other side of the barrier, provides an important tool for identification of partially unfolded forms. Furthermore, SX experiments reveal novel information about the role of side chain packing in folding that is not revealed by HX. Finally, by combining HX and SX experiments to access different kinetic regimes and probe backbone versus side chain exchange, we can create a more complete picture of the protein's energy landscape than either technique can provide alone.

5.5 References

1. Miyashita O, Onuchic JN, & Wolynes PG (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci U S A* 100(22):12570-12575.
2. Schrank TP, Bolen DW, & Hilser VJ (2009) Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. *Proc Natl Acad Sci U S A* 106(40):16984-16989.
3. Tripathi S & Portman JJ (2009) Inherent flexibility determines the transition mechanisms of the EF-hands of calmodulin. *Proc Natl Acad Sci U S A* 106(7):2104-2109.
4. Bemporad F, *et al*. (2008) Biological function in a non-native partially folded state of a protein. *EMBO J* 27(10):1525-1535.
5. Hyeon C & Onuchic JN (2007) Mechanical control of the directional stepping dynamics of the kinesin motor. *Proc Natl Acad Sci U S A* 104(44):17382-17387.
6. Hyeon C & Onuchic JN (2007) Internal strain regulates the nucleotide binding site of the kinesin leading head. *Proc Natl Acad Sci U S A* 104(7):2175-2180.
7. Terada TP, Sasai M, & Yomo T (2002) Conformational change of the actomyosin complex drives the multiple stepping movement. *Proc Natl Acad Sci U S A* 99(14):9202-9206.
8. Booth DR, *et al*. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* 385(6619):787-793.
9. Khurana R, *et al*. (2001) Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. *Biochemistry* 40(12):3525-3535.
10. Quintas A, Vaz DC, Cardoso I, Saraiva MJ, & Brito RM (2001) Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J Biol Chem* 276(29):27207-27213.
11. Bai Y (2006) Protein folding pathways studied by pulsed- and native-state hydrogen exchange. *Chem Rev* 106(5):1757-1768.
12. Berger A & Linderstrom-Lang K (1957) Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* 69:106-118.
13. Hvidt A & Nielsen SO (1966) Hydrogen exchange in proteins. *Adv Protein Chem* 21:287-386.
14. Bai Y, Sosnick TR, Mayne L, & Englander SW (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* 269(5221):192-197.
15. Bollen YJ, Kamphuis MB, & van Mierlo CP (2006) The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* 103(11):4095-4100.
16. Chamberlain AK, Handel TM, & Marqusee S (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* 3(9):782-787.
17. Fuentes EJ & Wand AJ (1998) Local dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* 37(11):3687-3698.
18. Krishna MM, Lin Y, Rumbley JN, & Englander SW (2003) Cooperative omega loops in cytochrome c: role in folding and function. *J Mol Biol* 331(1):29-36.
19. Maity H, Maity M, Krishna MM, Mayne L, & Englander SW (2005) Protein folding: the stepwise assembly of foldon units. *Proc Natl Acad Sci U S A* 102(13):4741-4746.

20. Vadrevu R, Wu Y, & Matthews CR (2008) NMR analysis of partially folded states and persistent structure in the alpha subunit of tryptophan synthase: implications for the equilibrium folding mechanism of a 29-kDa TIM barrel protein. *J Mol Biol* 377(1):294-306.
21. Yan S, Kennedy SD, & Koide S (2002) Thermodynamic and kinetic exploration of the energy landscape of Borrelia burgdorferi OspA by native-state hydrogen exchange. *J Mol Biol* 323(2):363-375.
22. Reiner A, Henklein P, & Kiefhaber T (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci U S A* 107(11):4955-4960.
23. Arrington CB & Robertson AD (1997) Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry* 36(29):8686-8691.
24. Bedard S, Mayne LC, Peterson RW, Wand AJ, & Englander SW (2008) The foldon substructure of staphylococcal nuclease. *J Mol Biol* 376(4):1142-1154.
25. Cliff MJ, Higgins LD, Sessions RB, Waltho JP, & Clarke AR (2004) Beyond the EX1 limit: probing the structure of high-energy states in protein unfolding. *J Mol Biol* 336(2):497-508.
26. Rodriguez HM, Robertson AD, & Gregoret LM (2002) Native state EX2 and EX1 hydrogen exchange of Escherichia coli CspA, a small beta-sheet protein. *Biochemistry* 41(7):2140-2148.
27. Sivaraman T, Arrington CB, & Robertson AD (2001) Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. *Nat Struct Biol* 8(4):331-333.
28. Cliff MJ*, et al*. (2006) A thiol labelling competition experiment as a probe for sidechain packing in the kinetic folding intermediate of N-PGK. *J Mol Biol* 364(4):810-823.
29. Feng Z, Butler MC, Alam SL, & Loh SN (2001) On the nature of conformational openings: native and unfolded-state hydrogen and thiol-disulfide exchange studies of ferric aquomyoglobin. *J Mol Biol* 314(1):153-166.
30. Ha JH & Loh SN (1998) Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat Struct Biol* 5(8):730-737.
31. Isom DG, Vardy E, Oas TG, & Hellinga HW (2010) Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity. *Proc Natl Acad Sci U S A* 107(11):4908-4913.
32. Jha SK & Udgaonkar JB (2007) Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* 282(52):37479-37491.
33. Johnson CP, Tang HY, Carag C, Speicher DW, & Discher DE (2007) Forced unfolding of proteins within cells. *Science* 317(5838):663-666.
34. Silverman JA & Harbury PB (2002) The equilibrium unfolding pathway of a (beta/alpha)8 barrel. *J Mol Biol* 324(5):1031-1040.
35. Sridevi K & Udgaonkar JB (2002) Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry* 41(5):1568-1578.
36. Krishnan B & Gierasch LM (2011) Dynamic local unfolding in the serpin alpha-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nat Struct Mol Biol* 18(2):222-226.

37. Goedken ER & Marqusee S (2001) Native-state energetics of a thermostabilized variant of ribonuclease HI. *J Mol Biol* 314(4):863-871.
38. Raschke TM & 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(4):298-304.
39. Silverman JA & Harbury PB (2002) Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J Biol Chem* 277(34):30968-30975.
40. Raschke TM, Kho J, & Marqusee S (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* 6(9):825-831.
41. Dabora JM & Marqusee S (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: characterization of a partially folded state. *Protein Sci* 3(9):1401-1408.
42. Yin J, *et al*. (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* 102(44):15815-15820.
43. Xiao H, *et al*. (2005) Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment. *Protein Sci* 14(2):543-557.
44. Goedken ER, Raschke TM, & Marqusee S (1997) Importance of the C-terminal helix to the stability and enzymatic activity of Escherichia coli ribonuclease H. *Biochemistry* 36(23):7256-7263.
45. Hollien J & Marqusee S (1999) Structural distribution of stability in a thermophilic enzyme. *Proc Natl Acad Sci U S A* 96(24):13674-13678.
46. Jha SK & Udgaonkar JB (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci U S A* 106(30):12289-12294.
47. Stratton MM, Cutler TA, Ha JH, & Loh SN (2010) Probing local structural fluctuations in myoglobin by size-dependent thiol-disulfide exchange. *Protein Sci* 19(8):1587-1594.
48. Santoro MM & Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27(21):8063-8068.

Chapter 6

Mutant tRNA-based translational misincorporation as a potential method to study ribosome-associated nascent chain structure

Experiments conducted in collaboration with Sabriya N. Rosemond, Kierstin L. Schmidt, and Tracy A. Young

6.1 Introduction

While *in vitro* protein folding studies have provided valuable insight into folding pathways and mechanisms, little is known about the first time any nascent chain folds, either as it is translated or once it is released from the ribosome. Both the vectorial nature of translation and possible interactions with the ribosome could make this process quite different from *in vitro* folding, or even folding of the full-length protein *in vivo*.

Detecting and characterizing structure in such complex systems, however, has been technically difficult. Recent studies have used various biochemical, biophysical, and computational methods, including NMR and fluorescence, to investigate nascent chain folding (1-10). Such studies have begun to shed light on the role of co-translational folding, but the picture remains incomplete, and the development of new methods to investigate this question is of paramount importance.

In this chapter, I will discuss the potential of an *in vivo* mutant tRNA-based misincorporation system to explore structure formation of the ribosome-associated nascent chain. Preliminary results with two homologous sets of proteins that show very different misincorporation behavior suggest that successful misincorporation could be correlated with the formation of a degradation-resistant structure on the ribosome, while failure to misincorporate reflects the lack of such structure.

Silverman *et al*. originally devised this *in vivo* misincorporation system, which uses mutant tRNAs to incorporate cysteines in the place of a given anticodon, to facilitate large-scale studies of protein structure and dynamics using cysteine-based methods (11). Briefly, the method works as follows: *E. coli* cells are transformed with two plasmids, one that codes for a protein of interest and another that codes for a cysteine tRNA with a mutated anticodon. This plasmid also contains a gene for the cysteinyl tRNA synthetase to ensure that the amount of synthetase in the cell is not the limiting factor for misincorporation. The misincorporator tRNA can then be charged with cysteine, which will be added to the nascent chain in place of the amino acid corresponding to the mutated anticodon. The misincorporator tRNA must compete with the naturally occurring native tRNA, so misincorporation is a relatively rare event, and most peptide chains have zero or one cysteine, resulting in a pool of mutants with single cysteine mutations, as well as some native protein.

This approach has shown high variability across the systems in which it has been attempted. The yeast triosephosphate isomerase (TIM) barrel was successfully misincorporated (12), as was the human TIP47 protein (13). The chaperone heat shock protein 90 (Hsp90), mouse syntaxin, and sea urchin calmodulin also show misincorporation (KLS, personal communication), though at relatively low levels for all proteins tested. There are other examples, however, where misincorporation has been extremely inefficient, to the point that it can be considered negligible. Misincorporation was attempted with three *E. coli* signaling proteins (CheA, CheW, and CheY); preliminary results indicate that CheA exhibited no detectable misincorporation levels, CheW showed variable and very low levels of misincorporation, and CheY showed detectable levels of misincorporation (Underbakke ES, personal communication).

This variation in misincorporation efficiency across many protein families from many different organisms suggests that there may be underlying fundamental reasons for misincorporation

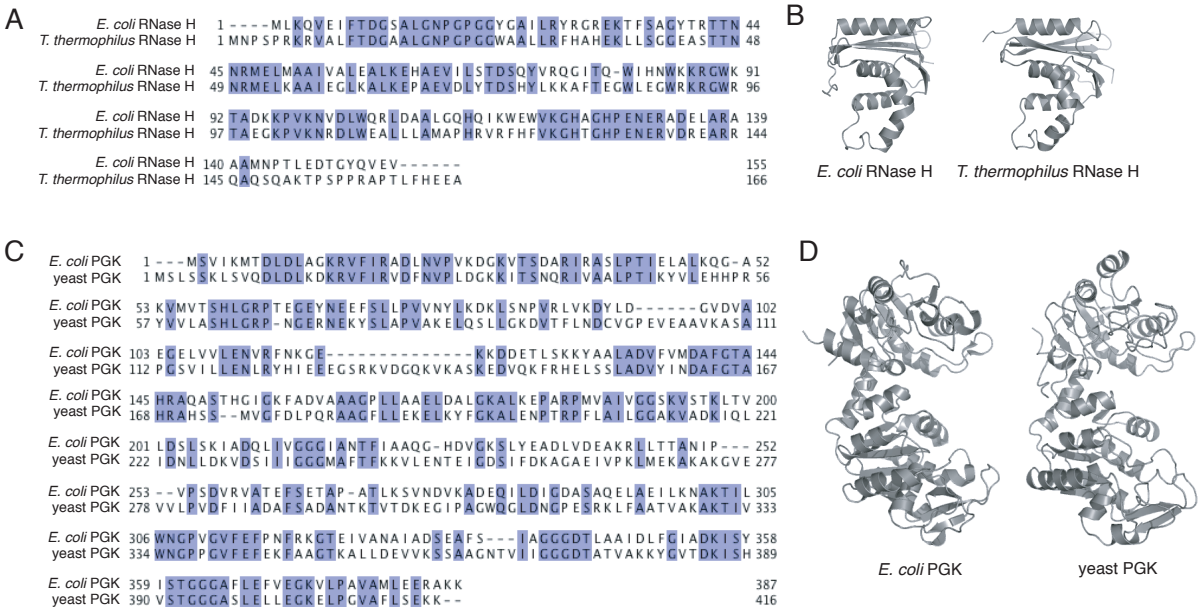
success or failure related to a protein's structure or translation dynamics. To investigate these effects, we applied the misincorporation method to two pairs of homologous proteins: *E. coli* and *Thermus thermophilus* ribonucleases H (RNase H), and *E. coli* and yeast phosphoglycerate kinases (PGK).

RNase H is a small, single-domain α/β protein that cleaves the RNA strand in DNA/RNA hybrids. The *E. coli* and *T. thermophilus* proteins have 52% sequence identity and 67% similarity and nearly identical tertiary structures (Figure 1A,B), although *T. thermophilus* RNase H is more stable at all temperatures (14). Both proteins populate folding intermediates that are believed to be similar in structure (15-17), but the folding kinetics for *T. thermophilus* RNase H are more complex than those for the *E. coli* protein (16). Furthermore, the two proteins have been found to have significantly different ΔC_p, which has been attributed to the presence of residual structure in the unfolded state of the protein from *T. thermophilus* (14).

PGK is a small glycolytic protein composed of two domains, each with a Rossmann fold topology, connected by an α-helical linker. The *E. coli* and yeast protein sequences are 39% identical and 56% similar, and the two proteins have highly similar tertiary structures (Figure 1C,D). Their thermodynamic stabilities are also similar, but the two proteins have different *in vitro* proteolytic susceptibilities, with *E. coli* PGK resistant while yeast PGK is not. Further studies showed that this difference arises from differing folding behavior of the domains: the isolated C-terminal domain of *E. coli* PGK cannot fold, while both domains of yeast PGK fold independently (18).

The combinations of shared and divergent biophysical characteristics in these two pairs of proteins provide a number of parameters to evaluate misincorporation. In this chapter I describe preliminary experiments that show distinctly different misincorporation behavior for the members of the two pairs. In both cases, the *E. coli* protein does not misincorporate well, while the non-*E. coli* protein shows relatively efficient misincorporation. Intriguingly, the E. coli proteins both lack some of the structure seen in the non-*E. coli* proteins – *E. coli* RNase H does not have the residual unfolded state structure suggested for *T. thermophilus* RNase H, and the isolated *E. coli* PGK C-terminal domain cannot fold, while that of yeast PGK can fold – suggesting that the misincorporation behavior may have a structural origin.

Figure 1. Amino acid sequence alignments and structures for the two pairs of homologous proteins. RNases H are shown in A and B; PGKs in C and D. Identical residues are highlighted in the sequence.



6.2 Methods

6.2.1 Gene construction

Cysteine-free variants of all four proteins (which have been previously reported ((14, 18, 19)) in all cases except yeast PGK, for which the one native cysteine was replaced with alanine) were cloned into a pET28-based vector with an N-terminal 6-His tag, a C-terminal ybbR tag for fluorescent labeling (20), and resistance to the antibiotic kanamycin. These constructs are called eRNH, tRNH, ePGK*, and yPGK*. The genetic constructs for both RNase H variants are synthetic genes similar to each other and largely unrelated to the organismal gene.

6.2.2 Misincorporation

Electrocompetent BL21(DE3) cells were co-transformed with two plasmids: the protein of interest (eRNH, tRNH, ePGK*, or yPGK*) and a mutant tRNA(11). Cells were plated on kanamycin (50 ug/mL) and carbenicillin (50 ug/mL) and incubated at 37°C overnight. A single colony was then picked and grown in LB to OD_{600} = 0.4-0.6. Cells were then spun at low speed for 10 minutes, washed twice in PBS, and used to inoculate M63 minimal media (30 ug/mL kanamycin and 50 ug/mL carbenicillin) in a 1:50 dilution. Cells were then grown to OD_{600} = 0.6, induced with 1 mM IPTG, and allowed to grow for an additional 14-16 hours, at which point they were harvested by centrifugation. This growth protocol proved quite challenging and often resulted in cell death; see appendix for further information.

6.2.3 Detection of misincorporation

Proteins were purified on a Ni-NTA column in the presence of 1 mM TCEP. Misincorporation was monitored by cleavage at positions of cysteine incorporation detected using a gel-based

fluorescence visualization method. Purified protein was incubated with fluorescein-CoA and the enzyme Sfp (20) for 15 minutes at room temperature, resulting in fluoresceine labeling at the C-terminal ybbR tag. Samples were subsequently incubated in 4 M GdmCl pH 8.6, 25 mM nitrothiocyanobenzoic acid (NTCB) at room temperature for 5 minutes, resulting in protein unfolding and cyanylation at cysteine residues. Protein was then collected for gel analysis by TCA precipitation, washed twice in acetone, and dried. To cleave protein at positions of misincorporation (cyanylation), pellets were resuspended in 10 ul 8 M Urea, 0.8 M sodium hydroxide and incubated for one hour at room temperature. After one hour, tubes were opened for about 20 minutes to allow ammonium hydroxide to evolve off (pH < 10). Samples were then run on a large, high-resolution gel using a modified Tris-Tricine system and visualized on a Typhoon fluorescence imager. Band intensities were quantified using ImageQuant.

6.3 Results

6.3.1 eRNH, but not tRNH, is refractory to misincorporation

Misincorporation was first attempted with eRNH. After many attempts, it was determined that misincorporation of this protein occurs at excessively low levels (Figure 2A). In all cases, overall expression was robust, but misincorporation occurred at very low levels, if detectable at all. In this system, both the mutant tRNAs and the protein of interest are under the control of the T7 promoter. To test if priming the cells with mutant tRNAs before inducing expression of eRNH might allow the misincorporation to proceed more successfully, the library of mutant tRNAs was put under the control of an arabinose-inducible promoter so they could be induced prior to protein expression. Using this system, tRNA expression was induced at various time points, up to two hours prior to induction of eRNH, but no increase in misincorporation was seen (KLS, personal communication). This result led us to return to the original T7-based system for all further experiments.

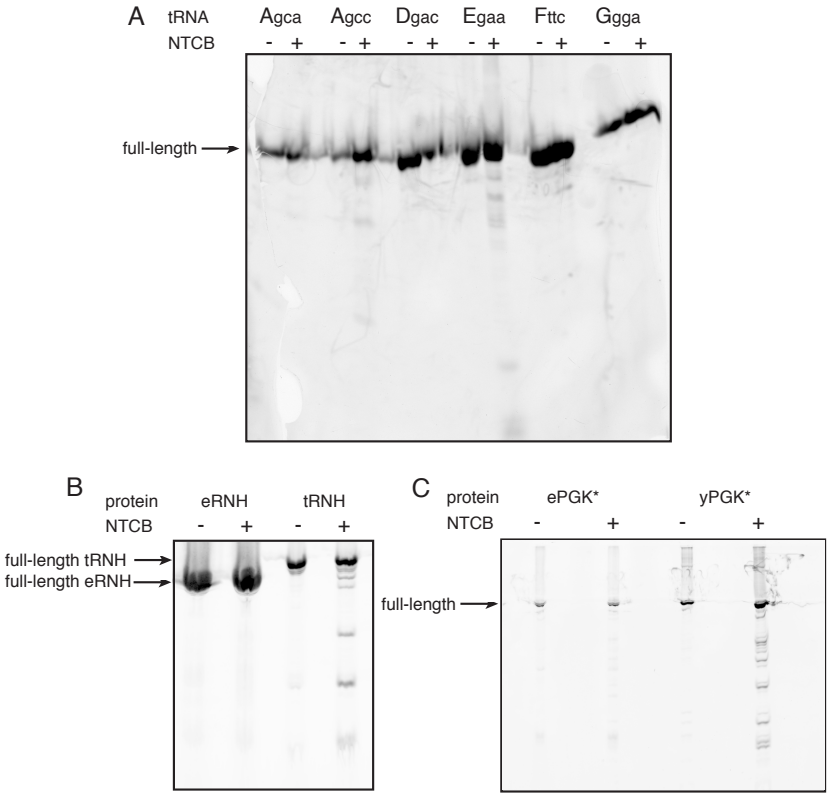
To test the protein-dependence of misincorporation, the phenylalanine misincorporator tRNA (tRNA-phe) was chosen because of its relatively high reported misincorporation efficiency in Hsp90 (KLS, personal communication). In contrast to the results for eRNH, tRNH showed robust misincorporation with this misincorporator (Figure 2B).

The simplest explanation for this discrepancy is that introducing cysteine mutations substantially destabilizes eRNH, making the misincorporated protein more highly susceptible to *in vivo* degradation or causing it to partition to the insoluble fraction during purification. Proteins with cysteine mutations introduced by QuikChange mutagenesis are expressed at high levels, however (see Chapters 4 and 5), suggesting that the difference is not simply due to the cysteine mutation.

6.3.2 yPGK, but not ePGK*, shows robust misincorporation*

Misincorporation experiments were carried out with tRNA-phe. yPGK* showed measurable misincorporation, but there was no detectable misincorporation for ePGK* (Figure 2C). It is possible that misincorporation for this protein is not seen because the expression levels were quite low; however, repeated attempts to produce more protein were not successful.

Figure 2. Misincorporation gels. Cleavage at cysteines should only occur in the presence of NTCB. (A) Misincorporation tests for eRNH with various mutant tRNAs. Extremely low levels of misincorporation are seen with some tRNAs (Agcc, Egaa, Fttc), but none show significant cleavage. (B) A comparison of the misincorporation efficiency for eRNH and tRNH using the Fttc misincorporator. Only tRNH shows measurable levels. (C) A comparison of the misincorporation efficiency for ePGK* and yPGK* using the Fttc misincorporator. Only yPGK* shows measurable levels.



6.4 Discussion

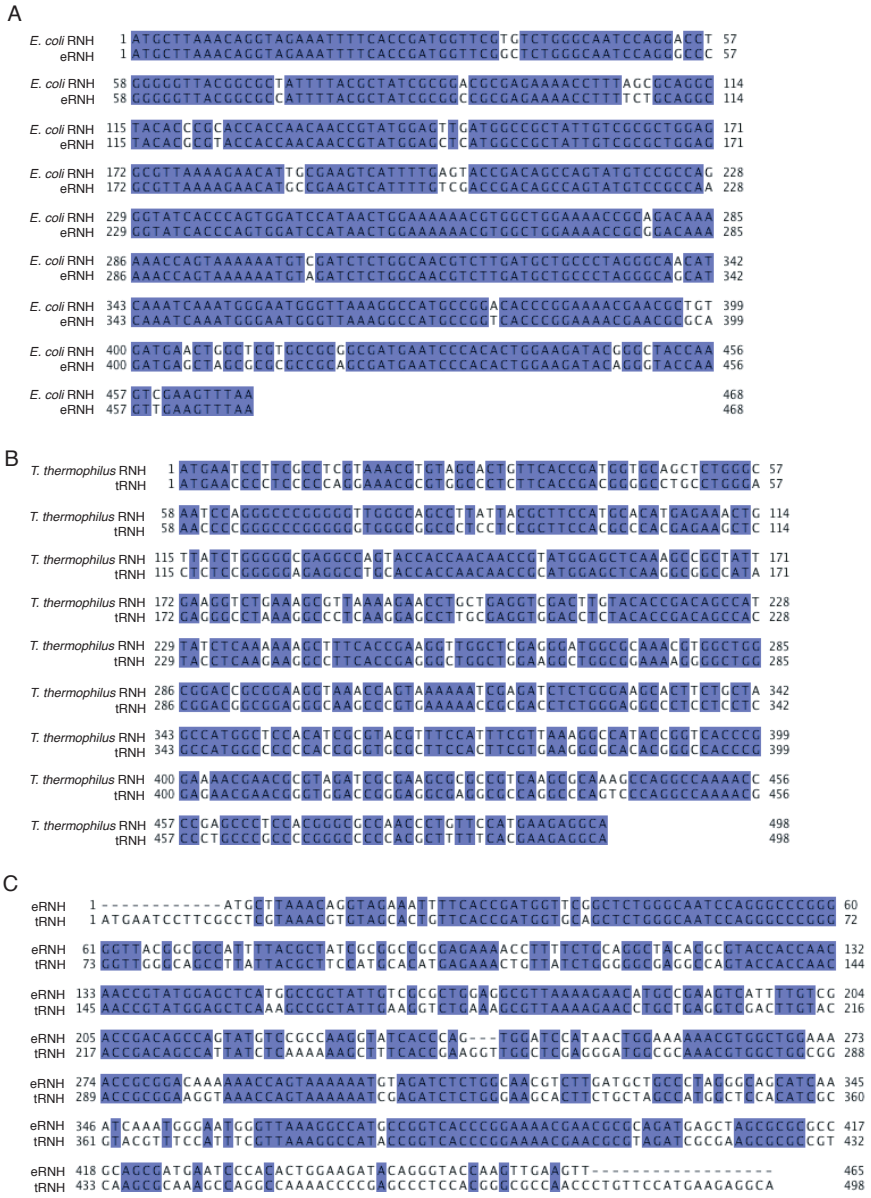
The results presented above show clear differences in misincorporation efficiencies for highly homologous proteins from different organisms with slightly different folding behavior. While there are a huge number of possible explanations for this behavior involving the incredible complexity of the translational machinery, there is some evidence to suggest that the results may be due to differential formation of ribosome-associated structure during translation.

6.4.1 Differences in misincorporation behavior seem to be based on protein, not DNA, sequence

The first hint that misincorporation behavior could act as a probe for ribosome-associated protein structure arose from the origins of the DNA sequences used for these experiments. Specifically, the *E. coli* RNase H gene used for these studies was synthesized to optimize its cloning sites and varies slightly from the endogenous *E. coli* RNase H sequence (Figure 3A). Furthermore, the *T. thermophilus* RNase H gene was constructed by introducing the minimal number of changes into the existing *E. coli* RNase H gene, resulting in a sequence in which almost a quarter of the bases deviate from the endogenous *T. thermophilus* RNase H DNA sequence (Figure 3B). Thus, the

resulting eRNH and tRNH DNA sequences used for misincorporation experiments are 66% identical despite only 52% identity in protein sequence (Figure 3C). The combination of the overwhelming similarity between the eRNH and tRNH DNA sequences and the nearly opposite misincorporation behavior suggested that misincorporation was more closely linked to the protein sequence than the DNA (or RNA) sequence.

Figure 3. DNA sequence alignments of experimental versus endogenous RNases H. Conserved residues in each pair are highlighted. (A) The endogenous gene coding for *E. coli* RNase H is 93% identical to that used for experiments (eRNH). (B) The endogenous gene coding for *T. thermophilus* RNase H is 76% identical that used for experiments (tRNH). (C) The eRNH and tRNH genes share 66% identity, despite only 52% identity in protein sequence.



6.4.2 Misincorporation may slow translation, leaving unstructured nascent chains susceptible to proteolysis or premature termination

Given the correlation between protein – and not DNA – sequence and misincorporation, it seems likely that misincorporation is related to protein structure. A likely model seems to be that that misincorporation is not seen for eRNH and ePGK because the misincorporated products are degraded as they are being produced due to altered translation kinetics upon misincorporation. Specifically, if the misincorporation event results in translational stalling, the nascent chain would be exposed, and if it does not gain significant structure during translation, it would likely be subject to proteolysis. Thus, perhaps eRNH and ePGK do not gain significant protective structure on the ribosome, resulting in their degradation upon translational stalling due to misincorporation, while tRNH and yPGK* fold into protected, partially folded structures and thus survive any misincorporation-induced stalling.

Such structural differences have been suggested by *in vitro* studies unrelated to the ribosome. The two RNases H have different ΔC_ps, which has been attributed to increased residual structure in the unfolded state for *T. thermophilus* RNase H* relative to *E. coli* (14). This result is suggestive that partially synthesized nascent chains for the two proteins could also show very different folding behavior on the ribosome. Studies of yeast and *E. coli* PGK showed that the isolated C-terminal domain of the *E. coli* protein cannot fold, while both the N- and C-terminal domains of the yeast protein independently fold into stable structures (18). This result also supports the hypothesis that the *E. coli* PGK nascent chain could have less structure on the ribosome before it is completely synthesized.

6.4.3 There are many possible mechanisms for misincorporation-induced translational stalling

It has been shown that naturally occurring tRNAs show nearly identical affinities to elongation factor Tu (EF-Tu) (21), rates of dissociated from the A and P sites of the ribosome (22), binding affinities to the ribosomal entry sites (23), rates of GTP hydrolysis (23), and peptide bond formation (23), and it has been suggested that this standardized behavior may have functional importance for translation. In fact, recent work showed that engineered valine tRNAs with altered EF-Tu binding affinities affect the rate of peptide bond formation (24).

The conservation of binding behavior for tRNAs with widely varying composition results from the binding interactions of a specific tRNA, with its cognate amino acid, to the mRNA and ribosome – a balance that could be upset by the misincorporation system, which attaches a cysteine amino acid to a tRNA with a non-cognate anticodon. The possibility that this energetic deviation from some established, optimized interaction might affect translation kinetics and cause slowing or stalling is an intriguing possibility.

Furthermore, recent results suggest that the identity of the amino acid conjugated to the tRNA can play a role in its ribosome binding and peptidyl transfer. The identity of the C-terminal amino acid in a nascent chain was shown to affect the rate of the peptidyl transfer reaction by an order of magnitude or more, independent of the identity of the tRNA (25). Although this study did not investigate cysteine, it is possible that, if misincorporation does in fact slow translation, it could partly be a result of simply having a cysteine in the P site. An additional study determined that native tRNAs synthetically conjugated to non-cognate amino acids show more non-productive binding events of the tRNA to the A site of the ribosome before a productive complex

is formed relative to the wild-type tRNA (26); a similar effect with the non-cognate misincorporator tRNAs could result in translational slowing upon misincorporation.

Translational editing by the ribosome could also cause the translational slowing required for our model. Recently, Zaher and Green suggested a ribosome-based editing mechanism that occurs after peptide bond formation to ensure translational fidelity (27). This mechanism relies on codon-anticodon recognition, and thus would not be implicated in our misincorporation system. It is not unreasonable, however, to suggest that there could be additional ribosome-based editing mechanisms that occur after peptide bond formation that might be activated by misincorporation. Such a system could result in slowed translation and thus increased degradation susceptibility for the unfolded nascent chains, while proteins that populate partially folded states on the ribosome would be protected from proteolysis. Furthermore, it has been shown that regions of the tRNA besides the anticodon are crucial for high fidelity peptide bond formation (28, 29), providing further evidence that such a non-anticodon-based editing system may exist.

6.4.4 Stalled nascent chains are degraded

If misincorporation does affect translation rates, there are a number of possible ways that a slowed or stalled ribosome-nascent chain complex could be degraded. Complexes that are stalled by rare codon clusters or tRNA depletion have been shown to be targets for ssrA-tagging, resulting in ribosome release and subsequent ClpXP-mediated degradation of the truncated polypeptide(30); it is possible that this mechanism is also acting on the misincorporated *E. coli* proteins, but not tRNH or yPGK*. A version of tRNH with a C-terminal ssrA tag was subject to degradation by ClpXP (31); however, when this same tag was attached to a version of eRNH, the *E. coli* protein was degraded much more quickly (JWD, personal communication). Furthermore, when the ssrA tag was appended to the thermophilic protein at internal positions in the protein via cross-linking at engineered cysteine residues, one attachment point resulted in significantly slowed degradation, and the other attachment point resulted in no degradation over the time-course of the experiment. Therefore, it seems possible that, in our misincorporation, slowed or stalled eRNH nascent chains are more susceptible to ssrA tagging and degradation than are tRNH.

Instead of complete degradation, it is also possible that the observed misincorporation results are due to differential premature termination within the protein pairs. Detection of misincorporation depends on the C-terminal ybbR tag for visualization, so any peptide chains missing this tag will not be seen. The mechanism by which misincorporation would cause termination in some cases but not others in unknown, but it also would likely depend upon misincorporation causing translation to slow or stall.

6.4.5 Slowing of translation affects folding

A number of recent studies have suggested that there is a strong connection between the rate of translation and cotranslational folding. Kimchi-Sarfaty *et al*. observed that a synonymous single-nucleotide polymorphism in the multidrug resistance 1 (MDR1) gene, which does not affect the amino acid sequence, results in altered activity for the protein; they hypothesize that this is a result of altered translational kinetics and thus altered folding (32, 33). Studies investigating the folding of eukaryotic proteins in *E. coli*-based heterologous expression systems shows that slowing translation promotes proper folding, which is attributed to favorable interactions with

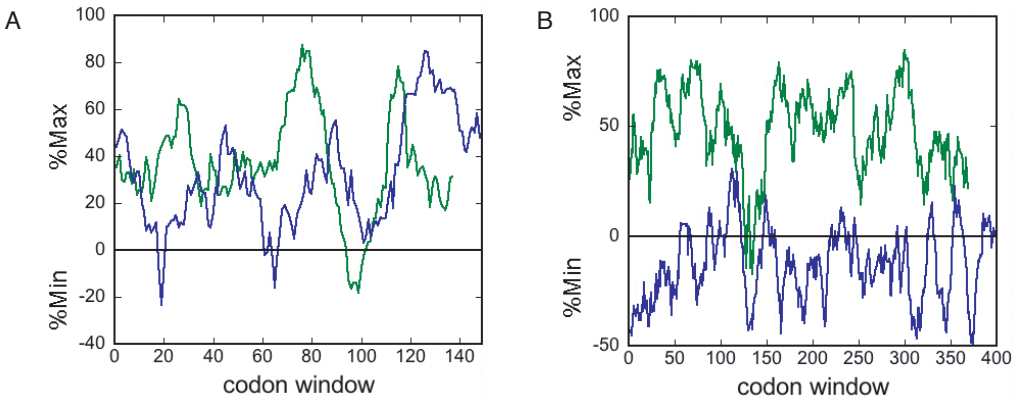
the ribosome for these proteins that, in their native organisms, generally show slower elongation rates than proteins in *E. coli* (34, 35). One can imagine that the opposite case could have an inverse affect: if misincorporation slows folding for *E. coli* proteins that are not optimized for interacting with and folding on the ribosome, their increased residence time as an unfolded nascent chain instead of a folded protein could leave them susceptible to proteolysis, detected as a lack of misincorporation.

6.4.6 Rare codon-related stalling analysis is consistent with model

Various experimental and computational studies suggest that clusters of rare codons in mRNA sequences can slow or stall translation (6, 36-38), providing an interesting analogy to the model for misincorporation-induced stalling. Specifically, if stalling by any mechanism results in proteolysis for some proteins, this effect could be detected by overall protein yield and correlated to misincorporation efficiency.

Analysis of the rare codon content of the proteins used for this study, however, is consistent with but do not confirm the misincorporation results. Calculations using the codon frequencies of *E. coli* reveal that there are no significant rare codon clusters in eRNH, tRNH, or ePGK*; yPGK*, on the other hand, has significant rare codon content (Figure 4A,B). The lack of rare codons in eRNH and ePGK* suggests that, in the absence of misincorporation, these genes should be synthesized at the normal rate; therefore, it is not unreasonable to suspect that misincorporation-based slowing – or slowing introduced by engineered rare codon clusters – could result in degradation. The fact that yPGK* is the only sequence with significant rare codon content and also shows misincorporation is also consistent with a differential degradation-based explanation for the misincorporation results.

Figure 4. Rare codon clusters in the four proteins subject to misincorporation, as calculated by the Clark rare codon calculator, using *E. coli* codon frequencies (37). Numbers below zero correspond to rare codons, while numbers above zero correspond to common codons. (A) Neither eRNH (green) nor tRNH (blue) show significant rare codon content. (B) ePGK* (green) does not have significant rare codon content; yPGK* (blue) is largely made up of rare codons.



6.5 Conclusions

Co-translational folding is an area of great interest, but studies in this field have generally been exceedingly complicated and difficult. While we have not confirmed that the differential misincorporation behavior seen here is directly related to ribosome-associated structure, there are a number of suggestive pieces of evidence. If we can confirm the relationship between misincorporation and co-translational folding, this relatively straight-forward *in vivo* system could be a highly valuable tool.

For proteins that have been studied in detail, misincorporation experiments with specific constructs could provide specific information about regions of the protein that are or are not folded during translation. Furthermore, it is possible that large-scale misincorporation experiments could provide a large amount of binary information about which proteins or types of proteins form structure co-translationally, providing a relatively high-throughput method to probe a question that up to this point has been exceedingly difficult to address.

With such studies, we may be able to begin to address questions regarding the *in vivo* relevance of partially folded states and kinetic intermediates detected by *in vitro* experiments. There has been some speculation that folded states on the ribosome may mimic such species, but no studies have yet been done to address this question. Given that *E. coli* and *T. thermophilus* RNase H are thought to have similar folding intermediates and equilibrium partially unfolded forms but suspected to have different structures in the unfolded state and clearly show different misincorporation behavior, it is tempting to say that structure on the ribosome is likely not directly correlated with such intermediates, but could be related to the structural propensity of the "unfolded" state.

6.6 References

1. Cabrita LD, Hsu ST, Launay H, Dobson CM, & Christodoulou J (2009) Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc Natl Acad Sci U S A* 106(52):22239-22244.
2. Eichmann C, Preissler S, Riek R, & Deuerling E (2010) Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc Natl Acad Sci U S A* 107(20):9111-9116.
3. Uemura S, *et al*. (2008) Single-molecule imaging of full protein synthesis by immobilized ribosomes. *Nucleic Acids Res* 36(12):e70.
4. Ellis JP, Bakke CK, Kirchdoerfer RN, Jungbauer LM, & Cavagnero S (2008) Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS chemical biology* 3(9):555-566.
5. Evans MS, Sander IM, & Clark PL (2008) Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J Mol Biol* 383(3):683-692.
6. Zhang G, Hubalewska M, & Ignatova Z (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 16(3):274-280.
7. Evans MS, Ugrinov KG, Frese MA, & Clark PL (2005) Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat Methods* 2(10):757-762.
8. Kadokura H & Beckwith J (2009) Detecting folding intermediates of a protein as it passes through the bacterial translocation channel. *Cell* 138(6):1164-1173.
9. Ziv G, Haran G, & Thirumalai D (2005) Ribosome exit tunnel can entropically stabilize alpha-helices. *Proc Natl Acad Sci U S A* 102(52):18956-18961.
10. Ugrinov KG & Clark PL (2010) Cotranslational folding increases GFP folding yield. *Biophys J* 98(7):1312-1320.
11. Silverman JA & Harbury PB (2002) Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J Biol Chem* 277(34):30968-30975.
12. Silverman JA & Harbury PB (2002) The equilibrium unfolding pathway of a (beta/alpha)8 barrel. *J Mol Biol* 324(5):1031-1040.
13. Burguete AS, Harbury PB, & Pfeffer SR (2004) In vitro selection and prediction of TIP47 protein-interaction interfaces. *Nat Methods* 1(1):55-60.
14. Hollien J & Marqusee S (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* 38(12):3831-3836.
15. Hollien J & Marqusee S (1999) Structural distribution of stability in a thermophilic enzyme. *Proc Natl Acad Sci U S A* 96(24):13674-13678.
16. Hollien J & Marqusee S (2002) Comparison of the folding processes of T. thermophilus and E. coli ribonucleases H. *J Mol Biol* 316(2):327-340.
17. Raschke TM & 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(4):298-304.
18. Young TA, Skordalakes E, & Marqusee S (2007) Comparison of proteolytic susceptibility in phosphoglycerate kinases from yeast and E. coli: modulation of conformational ensembles without altering structure or stability. *J Mol Biol* 368(5):1438-1447.

19. Dabora JM & Marqusee S (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: characterization of a partially folded state. *Protein Sci* 3(9):1401-1408.
20. Yin J, *et al*. (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* 102(44):15815-15820.
21. Louie A, Ribeiro NS, Reid BR, & Jurnak F (1984) Relative affinities of all Escherichia coli aminoacyl-tRNAs for elongation factor Tu-GTP. *J Biol Chem* 259(8):5010-5016.
22. Fahlman RP, Dale T, & Uhlenbeck OC (2004) Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol Cell* 16(5):799-805.
23. Ledoux S & Uhlenbeck OC (2008) Different aa-tRNAs are selected uniformly on the ribosome. *Mol Cell* 31(1):114-123.
24. Schrader JM, Chapman SJ, & Uhlenbeck OC (2011) Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. *Proc Natl Acad Sci U S A* 108(13):5215-5220.
25. Wohlgemuth I, Brenner S, Beringer M, & Rodnina MV (2008) Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *J Biol Chem* 283(47):32229-32235.
26. Effraim PR, *et al*. (2009) Natural amino acids do not require their native tRNAs for efficient selection by the ribosome. *Nat Chem Biol* 5(12):947-953.
27. Zaher HS & Green R (2009) Quality control by the ribosome following peptide bond formation. *Nature* 457(7226):161-166.
28. Cochella L & Green R (2005) An active role for tRNA in decoding beyond codon:anticodon pairing. *Science* 308(5725):1178-1180.
29. Ledoux S, Olejniczak M, & Uhlenbeck OC (2009) A sequence element that tunes Escherichia coli tRNA(Ala)(GGC) to ensure accurate decoding. *Nat Struct Mol Biol* 16(4):359-364.
30. Roche ED & Sauer RT (1999) SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *EMBO J* 18(16):4579-4589.
31. Kenniston JA, Burton RE, Siddiqui SM, Baker TA, & Sauer RT (2004) Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J Struct Biol* 146(1-2):130-140.
32. Kimchi-Sarfaty C, *et al*. (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 315(5811):525-528.
33. Tsai CJ, *et al*. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol* 383(2):281-291.
34. Siller E, DeZwaan DC, Anderson JF, Freeman BC, & Barral JM (2010) Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J Mol Biol* 396(5):1310-1318.
35. Agashe VR, *et al*. (2004) Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* 117(2):199-209.
36. Zhang G & Ignatova Z (2009) Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS One* 4(4):e5036.
37. Clarke TF & Clark PL (2008) Rare codons cluster. *PLoS One* 3(10):e3412.
38. Varenne S, Buc J, Lloubes R, & Lazdunski C (1984) Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* 180(3):549-576.

Appendix

Detailed materials and methods for misincorporation and native-state thiol alkyl-proton exchange

A1.1 Misincorporation

Electrocompetent BL21(DE3) cells (50 ul) were electroporated in a 1 mm gap electroporation cuvette at 1.7 V, 25 uFD capacitance, and 200 ohm resistance. 500 ul LB were added to cuvettes and cells were allowed to recover for 15 minutes at 37°C; 200 ul were then plated on LB plates with 50 ug/mL kanamycin and 50 ug/mL carbenicillin and incubated at 37°C overnight. A single colony was then picked and grown overnight; the overnight starter culture was then diluted into one quarter the desired final growth volume (usually 250 mL for a final growth volume of one liter). The culture was grown to OD_{600} = 0.4-0.6. Cells were then spun at low speed for 10 minutes, washed in PBS (equivalent to original growth volume), spun again at low speed for 10 minutes, and resuspended in equivalent amount of PBS. M63 minimal media (30 ug/mL kanamycin and 50 ug/mL carbenicillin) was then inoculated with resuspended cells in a 1:50 dilution (usually 4 mL cells to 200 mL media; attempting to grow larger volumes usually resulted in cell death). Cells were grown to OD_{600} = 0.4 and back-diluted 1:10 into additional M63 to obtain final desired growth volume. Cells were then grown to OD_{600} = 0.6, at which point they were induced with 1 mM IPTG and allowed to grow for an additional 14-16 hours, at which point they were harvested by centrifugation. (The 14-16 hours of expression has been optimized, but other growth times and volumes are flexible; the cells generally do not grow very well, especially in the M63 minimal media, so changes can be made depending on timing and convenience.)

M63 minimal media
50 mM KPi pH 7.5 (cannot use NaPi)
100 mM (NH4)2SO4
0.2% glucose
1 mM MgSO4
100 ul 0.5% thiamine HCl
1 crystal FeSO4
ddH20 to 1 L

PBS
20 mM KPi pH 7.5
150 mM NaCl

A1.2 Time-dependent alkylation

900 ul 11 uM protein samples (diluted to 10 uM upon addition of IAM) were prepared and allowed to equilibrate overnight in the appropriate [GdmCl] in 100 mM bicine pH 8.6, 50 mM KCl, 1 mM TCEP. Prior to alkylation, 90 ul of the protein sample was removed and added to an equal volume of quench buffer (2x labeling buffer plus βME) for the zero time-point. The alkylation reaction was initiated by the addition of 90 ul alkylation buffer (10x the final desired concentration of iodoacetamide in 100 mM bicine pH 8.6, 50 mM KCl). 100 ul aliquots were taken at varying time points and added to an equal volume of quenching buffer. Quenched samples were stored at -20°C until preparation.

To determine the kinetic exchange regime, [IAM] in the alkylation buffer was varied from 10 to 250 mM, resulting in final [IAM] from 1 to 25 mM. (IAM is only soluble up to about 500 mM in alkylation buffer.) The concentration of βME in the quench buffer was varied accordingly,

maintaining a 2.5x excess over the final [IAM]. The zero time-point samples were all treated with the standard 25 mM bME quench used for 10 mM IAM alkylation reactions.

A1.3 Alkylated sample preparation

Alkylated and quenched samples were fluorescently labeled by incubation for 15 minutes at room temperature with 2 ul fluorescein-CoA, 2 ul purified Sfp, and 20 ul 10x labeling buffer (0.5 M Tris pH 7.5, 0.1 M $MgCl_2$, 0.4% Triton). (If [GdmCl] in protein alkylation sample is above 1 M, additional dilution may be necessary for efficient labeling. Volumes of reagents for subsequent steps should be scaled up accordingly; care should be taken that the final volume upon completing the precipitation step will not exceed the volume of the tube used.)

Protein was subsequently unfolded and cyanylated by incubation at room temperature for 5 minutes with 200 ul (same as final volume after labeling) 8 M GdmCl, 200 mM bicine pH 8.6, and a final nitrothiocyanobenzoic acid (NTCB) concentration in a 2-fold excess over bME. (For example, if the original [IAM] used for alkylation was 10 mM, both [βME] and [NTCB] should be 25 mM.) A 1 M NTCB stock should be made in dry dioxane and stored at room temperature, and best if used fresh; reagent should not be used if it has strong orange color or is older than a week. The observed bright orange color upon addition of NTCB to quenched sample results from the reaction of the NTCB with free cysteines, resulting in the release of nitrothiobenzoate, analogous to the reaction of Ellman's reagent.

Protein was then collected for gel analysis by trichloroacetic acid (TCA) precipitation. 1 mL (five times volume after labeling) 0.05% sodium deoxycholate was added and mixed well, followed by the addition of 100 ul (half volume after labeling) 50% TCA. Upon mixing, the solution became cloudy as protein precipitated. The sample was then spun for 10 minutes at full speed in a microcentrifuge, pellet was washed twice in 600 ul acetone, and dried in a speedvac. (These steps are very important; one acetone wash is not sufficient to remove all excess NTCB, which will result in non-specific cleavage, and if the pellet is not completely dry before the next step, the cleavage efficiency will be very low).

Pellets were resuspended in 10 ul 8 M Urea, 0.8 M sodium hydroxide and incubated for one hour at room temperature. Care was taken to ensure that the small volume of urea was in appropriate contact with the protein pellet to solubilize and cleave. Tubes were occasionally vortexed to aid the process. After one hour, tubes were opened for about 20 minutes to allow ammonium hydroxide to evolve off (pH < 10).

Sfp growth and purification
BL21(DE3) cells transformed with the Sfp plasmid were grown in LB with kanamycin at 37°C, induced with 0.4 mM IPTG at OD_{600}=0.55, and harvested after 3 hours of expression.

The published protocol is as follows:
Resuspend cells in 30mL 10mM histidine pH6.5, 10% sucrose. Purify lysate in two column steps. The first step is a HiTrap Q Column equilibrated in 10mM histidine pH6.5, 2mM EDTA; protein is eluted with a gradient to 70% using the same buffer plus 500 mM NaCl at 2 mL/minute. Protein should elute between 15 and 20 minutes into gradient. Pool peak fractions and dialyze into 50mM Tris pH 8, 10mM $MgCl_2$, 5% glycerol, and concentrate to about 3 mL. Apply concentrated protein to a Sephacryl S-100 gel filtration column and run at 1 mL/minute; protein should elute between 125 and 160 minutes. Concentrate protein in same buffer plus 2.5 mM DTT.

I never had success with this protocol. While I was able to obtain a small amount of pure protein, I found that the flow-through from the Q column labeled ybbR-tagged protein much more efficiently than did the final, purified protein. To obtain larger amounts of more active protein, I used two ammonium sulfate precipitation steps; this resulted in relatively impure but enzymatically active protein. Cells were resuspended in 50 mL 50 mM Tris pH 8, 10 mM $MgCl_2$, 5% glycerol, and 5 mM DTT. The soluble portion of the lysate was subjected to a 35% ammonium sulfate precipitation with stirring for 1 hour at 4°C and spun at 12K RPM for 30 minutes. The soluble portion from this step is then subjected to a 55% ammonium sulfate precipitation following the same procedure. The insoluble portion from this precipitation step is resuspended in 50 mM Tris pH 8, 10 mM $MgCl_2$, 5% glycerol, and 5 mM DTT.

The resulting protein is relatively impure, but labels proteins well. Because of the difficulties purifying this protein, it is recommended that all fractions be retained and tested for their enzymatic activity. The impurities in the Sfp stock also help seed the protein pellet during the TCA precipitation. Therefore, if a more pure Sfp stock is prepared, 0.1 mg/mL BSA may have to be added to the quench buffer for quantitative TCA precipitation.

Synthesis of fluorescein-CoA
1.6 mg of coenzyme A in 1 mL 100 mM sodium phosphate pH7 was added to 1.3 mg of 5-iodoacetamidofluorescein in 0.5 mL DMSO and mixed for one hour at room temperature, wrapped in aluminum foil to protect from light. The reaction was purified over the reverse-phase C18 column on the HPLC with a gradient from 0 to 100% acetonitrile in 0.1% TFA over 60 minutes. Unreacted iodoacetamidofluorescein eluted slightly before fluorescein-CoA. Purity of fluorescein-CoA was tested by running a test labeling reaction; iodoacetamidofluorescein labeled all proteins nonspecifically, while fluorescein-CoA labeled only those proteins with the ybbR tag (catalyzed by Sfp). Fluorescein-CoA was stored at -80°C.

Labeling buffer
50 mM Tris pH 7.5
10 mM $MgCl_2$
0.04% Triton

A1.4 Sample visualization by gel electrophoresis
Samples in urea and ammonium hydroxide were combined with 5 ul 3x loading dye, and 10 ul of sample/well was loaded into each gel. Gels were run using a discontinuous, tris-tricine-based protocol in the cold room at at 700 V, 75 mA, 40W (increase current to 150 mA if two apparatuses are connected to power supply). Fluorescein emission was detected on a Typhoon imager by excitation with a 520 nm-bandpass (520 BP 40) laser.

3x sample loading dye
0.5 M Tris pH 6.5
25% glycerol
5% SDS

Modified tris-tricine gels (recipe provides enough for two large gels)

	Comb	Stacking	Intercalating	Resolving
40% 19:1 acrylamide	2.0 mL	1.0 mL	2.6 mL	16.0 mL
Gel buffer	-	2.4 mL	3.4 mL	12.0 mL
Comb buffer	3.4 mL	-	-	-
50% glycerol	-	-	2.0 mL	7.2 mL
dH20	4.6 mL	6.6 mL	2.0 mL	0.8 mL
total	10.0 mL	10.0 mL	10.0 mL	36.0 mL
Add just before pouring:				
10% APS (fresh)	100 ul	100 ul	100 ul	200 ul
TEMED	14 ul	14 ul	14 ul	40 ul

Gel buffer
3 M Tris pH 8.6
0.3% g SDS

Comb buffer
0.3 M Tris base (un-pHed)
0.3 M tricine
0.3% SDS

Anode (bottom) running buffer
0.2 M Tris pH 8.8

Cathode (top) running buffer
0.1 M Tris base (un-pHed)
0.1 M Tricine
1% SDS