Characterizing Rare and Transient Conformations of Proteins using Amide Hydrogen Exchange and Thiol Exchange

By

Eric R Bolin

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

In

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Susan Marqusee, Chair
Professor Andreas Martin
Professor David Wemmer
Professor Gloria Brar

Summer 2018
Abstract
Characterizing rare and transient conformations of proteins using amide hydrogen exchange and thiol exchange

by Eric R Bolin
Doctor of Philosophy in Biophysics
University of California, Berkeley
Professor Susan Marqusee, Chair

The dynamic nature of proteins is often an underappreciated aspect of biology necessary for understanding protein function and turnover in the cell. This is in part because within the multitude of states that a protein occupies throughout its lifetime, at most a few are significantly populated and have long enough lifetimes to study using traditional structural biology methods. In this thesis, I develop and explore experimental methods (amide hydrogen-deuterium exchange and cysteine labeling) to selectively label and characterize these rare and transient conformations methods. Specifically, I used cysteine labeling as a means to validate computational predictions of rare conformational fluctuations that expose potential drug-binding sites, I used pulse-labeling amide hydrogen exchange coupled with mass spectrometry detection to identify intermediates formed during protein folding and developed new data analysis procedures for analyzing hydrogen exchange rates measured determined by mass spectrometry on proteolytic fragments.

The first project used thiol labeling rates to validate computational predictions of cryptic binding sites. Analysis of millisecond-long molecular dynamics (MD) simulations of β-lactamase uncovered several potential conformational fluctuations that expose hidden, or cryptic binding pockets. Targeting these rare conformations presents a potential avenue for drug development. I validated the existence of these pockets by introducing cysteine residues in select positions and characterizing their accessibility to chemical modifications. In addition to validating the existence of these fluctuations, modification of cysteines at these sites modulates activity allosterically.

In the second project, I used pulsed labeling hydrogen exchange to follow the folding pathway of a protein family over evolutionary time (>3 billion years). I determined the conformations populated during folding (folding intermediates) for a family of RNase H proteins, including two extant and seven ancestral proteins. Each of these proteins was shown to populate a similar folding intermediate prior to the rate-limiting step in folding; however, the details of the steps leading up to this intermediate varied. We further showed that we can alter these early folding steps for a given protein using rationally designed mutations.

The third project compared and characterized the refolding and co-translational folding pathway of the protein HaloTag. We found that HaloTag aggregates during refolding but not during cotranslational folding, and it adopts at least one intermediate during refolding that is suppressed during cotranslational folding.
Finally, in the fourth project I developed a modification of a recently published approach for analyzing data from mass spectrometry-based detection of hydrogen exchange. This modification allows me to obtain multiple protection factors per peptide monitored and obtain a quantitative measurement of the free energy of hydrogen exchange in different regions of the protein.
Table of Contents

Abstract 1
Table of Contents i
List of Figures iii
Acknowledgements v

Chapter 1: Introduction 1
1.1 Introduction 2
1.2 Protein Structure and Folding – The Energy Landscape 2
1.3 Labeling High-Energy States 4
1.4 Cryptic Allosteric Sites 7
1.5 Folding Pathways and Intermediates 7
1.6 Energy Landscapes of RNases H 8
1.7 Cotranslational Folding through the lens of HaloTag 10
1.8 Mass Spectrometry and Data Analysis 10
1.9 Summary 11
1.10 References 12

Chapter 2: Identifying cryptic binding sites: combining excited-state labeling with computation to reveal novel drug binding sites 18
2.1 Abstract 19
2.2 Introduction 19
2.3 Results and Discussion 22
2.4 Conclusions 29
2.5 Materials and Methods 29
2.6 References 32
2.7 Supplemental Materials 35

Chapter 3: Evolutionary examination of folding intermediate structure leads to tunable folding pathways 40
3.1 Abstract 41
3.2 Introduction 41
3.3 Results 43
3.4 Discussion 50
3.5 Materials and Methods 53
3.6 References 55
3.7 Supplemental Materials 60

Chapter 4: Formation of an aggregation-prone folding intermediate that is suppressed during translation. 68
4.1 Abstract 69
4.2 Introduction 69
4.3 Results 70
4.4 Discussion 78
4.5 Materials and Methods 78
4.6 References 84
4.7 Supplemental Materials 87
# Chapter 5: Increased sequence resolution of hydrogen exchange mass spectrometry studies through sub-peptide level protection factors

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Abstract</td>
<td>98</td>
</tr>
<tr>
<td>5.2 Introduction</td>
<td>98</td>
</tr>
<tr>
<td>5.3 Method</td>
<td>99</td>
</tr>
<tr>
<td>5.4 Results</td>
<td>101</td>
</tr>
<tr>
<td>5.5 Discussion</td>
<td>104</td>
</tr>
<tr>
<td>5.6 References</td>
<td>104</td>
</tr>
<tr>
<td>5.7 Supplemental Information</td>
<td>105</td>
</tr>
</tbody>
</table>
List of Figures

Chapter 1

| Figure 1-1 | Protein Folding Funnel | 3 |
| Figure 1-2 | Labeling of Nonnative states | 6 |
| Figure 1-3 | The folding pathway of ecRNH* | 9 |

Chapter 2

| Figure 2-1 | Crystal Structure of TEM-1 β-lactamase | 20 |
| Figure 2-2 | Thiol labeling of the known hidden allosteric site | 22 |
| Figure 2-3 | Labeling and Unfolding of L286C | 25 |
| Figure 2-4 | Residues selected for labeling in each pocket | 27 |
| Figure 2-5 | Thiol labeling of the first predicted hidden allosteric site | 28 |
| Figure 2-S1 | Residuals for the single exponential fit to Fig. 2C | 38 |
| Figure 2-S2 | Thiol labeling of the second predicted hidden allosteric site | 38 |
| Table 2-1 | Specific activities of labeled and unlabeled proteins | 26 |
| Table 2-S1 | Experimental unfolding rates monitored by CD and DTNB labeling rates | 39 |

Chapter 3

| Figure 3-1 | RNase H structure and Pulsed-labeling HX-MS | 42 |
| Figure 3-2 | Determination of the folding pathway of T. thermophilus RNase H* by HX-MS | 45 |
| Figure 3-3 | Determination of the folding pathway of ancestral RNases H by HX-MS | 47 |
| Figure 3-4 | Intrinsic helicity as a predictor for the early folding mechanism of RNases H | 48 |
| Figure 3-5 | Engineered mutations to alter the folding pathway of ecRNH* | 50 |
| Figure 3-S1 | Chevron plot of RNase H variants studied at 10°C and 25°C | 60 |
| Figure 3-S2 | Determination of the folding pathway of Anc2* by HX-MS | 61 |
| Figure 3-S3 | Determination of the folding pathway of Anc3* by HX-MS | 62 |
| Figure 3-S4 | Determination of the folding pathway of AncA* by HX-MS | 63 |
| Figure 3-S5 | Determination of the folding pathway of AncB* by HX-MS | 64 |
| Figure 3-S6 | Determination of the folding pathway of AncC* by HX-MS | 65 |
| Figure 3-S7 | Determination of the folding pathway of AncD* by HX-MS | 66 |
| Table 3-S1 | Comparison of intrinsic helicity and AABUF across RNase H variants | 67 |

Chapter 4

| Figure 4-1 | Cotranslational folding of HaloTag monitored in real-time with FP | 71 |
| Figure 4-2 | Characterization of HaloTag folding kinetics and stability | 73 |
| Figure 4-3 | HaloTag folding is more efficient during in vitro translation than after refolding | 75 |
| Figure 4-4 | The HaloTag folding trajectory changes during cotranslational folding | 77 |
| Figure 4-S1 | Cotranslational folding of HaloTag can be measured using FP | 87 |
| Figure 4-S2 | Addition of the peptidyl-proline isomerase CypA does not affect HaloTag refolding or cotranslational folding rates | 88 |
| Figure 4-S3 | Aggregation of HaloTag | 89 |
| Figure 4-S4 | Cysteine accessibility of WT HaloTag | 89 |
| Figure 4-S5 | Characterization of Halo* cysteine mutants | 90 |
**Figure 4-S6** Gels for data shown in Fig. 4

**Figure 4-S7** Characterization of Halo* E121C cysteine accessibility

**Table 4-1** Summary of kinetic and thermodynamic data of HaloTag and its mutants

**Table 4-S1** Kinetic data obtained for HaloTag folding using FP

**Table 4-S2** Determination of HaloTag folding efficiency under different conditions

**Table 4-S3** Folding rates of HaloTag and variants measured by FP

**Table 4-S4** Normalized HX-MS data

**Chapter 5**

**Figure 5-1** Analysis of Peptide Deuteration

**Figure 5-2** Error in fitting PFs

**Figure 5-3** Comparison of TR fitting and exponential fitting

**Figure 5-4** Comparison of Mass Spec and NMR detection

**Figure 5-S1** Rate fitting accuracy

**Figure 5-S2** Accuracy of fitting of multiple PFs per peptide

**Figure 5-S3** Single Protection Factor Fitting

**Figure 5-S4** Plot of TR values

**Table 5-S1** Per-residue Protection Factors for ecRNH*
Acknowledgements

It goes without saying that the work in this thesis could not have been completed without the support of many people. The contributions of my advisor, Susan Marqusee are too numerous to count. Susan, I thank you for all of the roles you've fulfilled in my time in grad school: mentor, advisor, advocate, and ally. In addition, this work could never have happened without such a knowledgeable, supportive, fun lab. I thank all the members of the Marqusee lab for being there and helping throughout. I especially thank Shion Lim and Avi Samelson whose collaborations and scientific guidance were critical to this thesis. I also thank Helen Hobbs for all the coffee breaks that were indispensable in the completion of this work. In addition, there hasn't been a person in the lab who hasn't played some role in supporting me and this work, and I cannot thank them enough.

Outside of lab, I have had an amazingly supportive group of friends and family. I cannot thank my family enough for their support in all that I do. I thank my friends Jamie Lincoff, Mark Shapero, and Patrick Wilson for helping preserve whatever sanity I have left throughout this process. And finally, I cannot thank Keith Cheveralls enough for being a supportive partner throughout every struggle and adversity represented by this work. This thesis and the time I spent completing it would never have been possible without any of you.
Chapter 1:
Introduction
1.1 Introduction

Proteins are present in all cells and are responsible for carrying out most of the biological reactions and functions required for life. For many years, the study of proteins was driven by the dogma that ‘structure determines function’ and focused on determining the static three-dimensional structure of proteins. Over the last several decades, however, it has become clear that proteins are dynamic molecules that also occupy rare and transient alternative conformations and that these conformations and their access from the native conformation, called fluctuations, are crucial for protein function. While there are well-developed approaches for determining the three-dimensional structure of proteins, identifying and characterizing the structures sampled during these dynamic excursions presents a significant challenge as these conformations are often transient, rare, and difficult to detect.

How does one characterize a structure that may exist in less than one in a million molecules and may only be present for milliseconds? One powerful approach is to use a molecular probe that will selectively label proteins as they sample these so-called excited states. This approach has several advantages. One is that it allows for the detection of the labeled molecules after they have exited an excited conformation, enabling the observation of extremely transient states. Another advantage relies on its selectivity – allowing one to label these rare conformations, despite the presence of many other states that are more populated.

The work presented here focuses on the development and application of two such approaches: thiol labeling and amide hydrogen exchange. These methods are used to determine novel protein fluctuations that expose drug binding sites in the protein β-lactamase, identify conformations that members of the RNase H family of proteins sample as they search for and find their native folded structure, examine how the folding pathway of a protein can be influenced by translation on the ribosome using the model protein HaloTag, and determine protection factors for the quantitative interpretation of mass spectrometry data using newly developed methods.

1.2 Protein Structure and Folding – The Energy Landscape

Proteins are composed of amino acids. There are 20 naturally occurring amino acids and typical proteins are composed of hundreds or even thousands of amino acids linked together in a particular sequence to form a polypeptide chain. This means that there is an extremely large number of possible protein sequences, and proteins can contain one or several specific polypeptide chains. In 1973, Christian Anfinsen observed that the protein ribonuclease A was able to spontaneously adopt its active structure upon dilution from denaturing conditions. Thus, this amino acid sequence dictates a protein’s fold and proteins are capable of finding this structure on their own under certain conditions.1

While extremely important, this three-dimensional structure, usually termed the native or folded state, is not sufficient to explain the biological function and turnover of many proteins. For instance, proteins must be dynamic in order to function as enzyme catalysts. Crystalline ribonuclease A is known to lose function at temperatures that restrict dynamics,2 the small rotameric motions and conformations that cyclophilin A populates have been shown to be critical to its catalytic function,3 and recent de novo designed enzymes are often poor catalysts due to our poor understanding of the
dynamics that drive function in natural enzymes.⁴⁻⁶ At an even more fundamental level, the folded state of a protein is also insufficient to fully explain both protein folding and protein degradation, since both of these processes require dynamic access to non-native conformations. Thus, nearly all critical aspects of protein function require access to more than a single, unique structure.

One prominent example of how high-energy conformations can be critical in understanding function is found by comparing two kinases, Abl and Src. Although these proteins are 54% identical in their amino acid sequence and have almost identical structures, the drug imatinib is known to preferentially bind to Abl with 3000 times the affinity of its binding to Src. This difference in affinity is predominantly the result of a difference in relative populations of specific non-native high-energy states as well as the dynamics of interconversion between these states.⁷ This highlights that a thorough understanding of the high-energy states a protein adopts under native conditions is essential to understanding protein function even when the three-dimensional structure is known.

A protein will sample many such high-energy conformations even under native conditions (the set of conditions where a protein is usually functional) despite the fact that the vast majority of molecules are in the native structure. The population of each accessible conformation will be proportional to its Boltzmann factor. The rate of interconversion between these conformations is determined by the barriers between them. This complex ensemble can be modeled as an energy landscape that is often depicted as a ‘folding funnel’. In this model (see Fig.1-1), under native conditions, the native state is represented as the bottom of the funnel or the deepest “well” while other potential conformations are represented as nearby minima with a depth corresponding to their relative stability. The top, or ridge, of the funnel represents the diverse array of high-energy unfolded or random conformers.⁸⁻⁹ In this view, the z-axis represents internal energy of the protein while the xy-plane depicts conformational entropy.

![Figure 1-1. Protein Folding Funnel](image)
Adapted from Dill and MacCallum (2012). The folding funnel of a protein represents the energy landscape for a protein under native conditions. The xy-plane represents entropy while the z-axis reflects the internal energy of each protein conformation. The native state is the major well in the center of the depiction with random coil conformations represented as a flat plane near the periphery.
This landscape is particularly useful for thinking about protein folding, the process by which an unfolded or denatured protein navigates its way to the native, or ground state. Protein folding can be thought of as a journey starting at the top of the funnel in a high-energy unfolded state and moving down to the bottom of the funnel. Along the way, the protein loses both conformational freedom (that is, entropy) and internal energy. As it traverses down this funnel, a protein might “slide” smoothly and directly to the native state, or it might encounter a shallow well and populate one or several non-native structure(s) transiently. These shallow wells encountered during folding are termed ‘folding intermediates’ and are one important type of high-energy state that a protein can sample. Folding intermediates are discussed at length later in this introduction.

1.3 Labeling High-Energy States

It is clear that identifying these high-energy conformations is critical to understanding both protein function and folding. Characterizing these states, however, is a significant technical challenge, because, by definition, the native state is the predominant structure under native conditions. Thus, the native state will dominate any spectroscopic signal used to study protein structure. One successful solution to this problem is to study a single molecule as it traverses its landscape. Signal from non-native states is not masked by the large population of native state structure allowing these conformations to be examined. While such single molecule studies have successfully identified the presence and structural details of high-energy protein conformations, they are still limited by the fact that these high energy conformations are rarely sampled and transiently populated.

An alternative approach is to selectively label the high-energy state in order to generate a specific signal that is unpolluted by the more populated native state. For instance, using selective chemistry, one can label functional groups that are buried or inaccessible in the native state but become accessible during these fluctuations into non-native conformations. There are many types of chemistries that can be harnessed for these labeling strategies. The work presented in this thesis uses two such chemistries. The first, called amide-hydrogen exchange, isotopically labels the protein backbone. In this case, the native state is usually inaccessible to exchange due to hydrogen bonding or burial from the solvent. However, in a non-native conformation, the same amide hydrogen may be solvent-accessible. The second labeling strategy used in this work is an approach called thiol exchange in which solvent accessibility of a cysteine side-chain thiol is monitored. Both of these methods can be used to obtain information on the structure, thermodynamics, or kinetics of these high-energy states.

Theory of Exchange

Both amide-hydrogen exchange and thiol exchange harness the fact that the rate of labeling is affected by the accessibility of the functional group. In other words, sites that are less accessible due to protein structure will be labeled more slowly than sites that are unstructured. This slowing is what is monitored in a hydrogen- or thiol-exchange study. The ‘amount’ of slowing is quantified by the ratio of the expected labeling rate of an unstructured molecule to that of the observed rate. This ratio is usually referred to as the protection factor, or P. In native proteins, protection factors.
can vary from one for unstructured regions of a protein to more than a million for stable structures and can be used to characterize the thermodynamics and/or kinetics of transition between the ground state and the accessible conformation that gets labeled.

The theory underlying these protection factors was first proposed by Kaj Ulrik Linderstrøm-Lang. The model proposed by Linderstrøm-Lang is diagramed in Figure 1-2. In brief, each labeling site can exist in either a closed and labeling-incompetent state or in an open and labeling-competent state. These two states exist in equilibrium with an opening equilibrium constant $K_{op}$ and rates of interconversion $k_{op}$ and $k_{cl}$. Labeling can occur only from the open state with rate $k_{int}$ and is considered irreversible. While in the case of thiol exchange this process is actually irreversible, in the case of hydrogen exchange the experiment is usually set up such that after the first labeling reaction any future exchange is undetected. This is accomplished by dissolving the protein in a solution of a different isotope (H to D or D to H) and therefore any subsequent exchanges with the solvent do not involve changing the isotopic label.

For any protected sites in the native protein, $k_{op} \ll k_{cl}$ and the majority of molecules exist in the closed state. In this condition, the overall rate of labeling, $k_{obs}$ is related to the intrinsic rate and all the other opening and closing rates by

$$k_{obs} = \frac{k_{op} * k_{int}}{k_{op} + k_{cl} + k_{int}}$$

From this expression, it is clear that, in general, it is difficult to extract kinetic or thermodynamic properties from the observed rate. However, there are two limiting cases where we can obtain these data. In one extreme, abbreviated EX1, where $k_{int} >> k_{cl}$, every time a labeling site becomes exposed to solvent, it gets labeled. Thus $k_{obs} = k_{op}$, and the labeling rate directly measures the rate of exposure (kinetics) of this site. In another extreme, abbreviated EX2, $k_{int} << k_{cl}$, and the open and closed states exist in an equilibrium prior to labeling. The observed labeling rate is thus a measure of this equilibrium population; specifically, the observed rate is the intrinsic rate of labeling multiplied by the equilibrium constant (thermodynamics) for opening, $k_{obs} = k_{int} * K_{op}$. Since $k_{int}$ can be either easily measured experimentally or calculated, in this regime $K_{op}$ can be obtained.

Whether a particular site falls into one of these regimes depends on its intrinsic rate of exchange relative to the kinetics of the closing reaction. Fortunately, the intrinsic rates of both thiol and hydrogen exchange can be experimentally manipulated. In hydrogen exchange, which is catalyzed by both acid and base, $k_{int}$ can be predictably changed by changing the pH of the solution. In thiol exchange, the exchange rate will depend on the concentration of label and the chemistry of the labeling reagent. Thus both approaches can be used to measure thermodynamic and kinetic parameters.

This flexibility lends itself to several types of experiments to measure stability or dynamics under a range of conditions. Continuous labeling experiments can measure thermodynamic and kinetic parameters of high-energy states under native conditions. In these experiments, a protein is incubated with label under native conditions and labeling occurs when the protein samples high-energy states in which the labeling site is exposed to the solvent during excursions from the native state. Pulsed-labeling
experiments can be used to monitor protein structure during kinetic processes such as protein folding. In this case, the kinetic process is initiated and then after a delay time, a ‘pulse’ of label is applied to capture a snapshot in time of the protein’s structure. To monitor protein folding, refolding is typically initiated by dilution of a chemical unfolding agent, and a short pulse of label is used to label any site that is not stably folded. By varying refolding times, the individual conformations a protein adopts throughout this process can be followed with exquisite temporal and structural resolution.

### Thiol Exchange

Thiol exchange is capable of measuring the exposure of buried side chains in the protein. Cysteine side chains have a thiol moiety that can be labeled covalently with thiol exchange or maleimide chemistry. Natural or engineered cysteines that are buried in the native folded structure can thus be used as probes for tertiary structure and side-chain packing and burial in high-energy conformations. Labeling with molecules of varying sizes can be used to determine the size or shape of openings present in these conformations as well as the kinetics and thermodynamics of the openings. In addition, determining whether a site falls into the EX1 or EX2 regimes is simple, as the dependence of $k_{obs}$ on the labeling molecule concentration can distinguish between EX1 (no dependence) from EX2 (linear dependence) directly.

### Hydrogen Exchange

Unlike thiol exchange, which can typically only measure exposure of a single residue at a time, hydrogen exchange experiments probe every amide hydrogen along the entire protein backbone. Each peptide bond in the amino acid chain (with the exception of those linking to the proline amine) has an amide hydrogen capable of exchange. Protection of these hydrogens indicates the formation of protein secondary
structure or hydrogen bonding at this site. This makes hydrogen exchange a powerful technique to observe either stability or dynamics across an entire protein sequence. Once again, this technique can be tuned to measure either kinetics (the EX1 regime) or thermodynamics (the EX2 regime) by changing in the pH of the solution. Since the exchange reaction is both acid- and base-catalyzed, adjusting the pH of the solution changes the $k_{\text{ex}}$ by a factor of ten per pH unit. There is a minimum in the intrinsic exchange rate around pH 3; the rate increases as the pH is adjusted away from this value in either direction.$^{24,25}$ Thus, adjusting the pH of the solution can push the protein to exchange in either regime and give information on both kinetic and thermodynamic parameters under similar conditions with the caveat that pH can alter stability for certain systems.

### 1.4 Cryptic Allosteric Sites

One area in which observing high-energy conformations could yield novel insights is the study of cryptic drug binding sites. These sites are defined as ligand-binding pockets on a protein that are “hidden” or undetectable in the known structures of a protein. Despite the apparent absence of these pockets in the native conformation of the protein, binding to these pockets can allosterically modulate protein activity,$^{34}$ and such sites have been identified during drug design without any prior knowledge of the pocket.$^{35}$ These sites offer the exciting possibility of developing drugs targeting proteins that were previously thought to be undruggable. However, the general utility of these sites depends on the ability to characterize them prior to undertaking expensive drug design studies. Unfortunately, predicting the presence of these hidden allosteric sites de novo is difficult as they represent high-energy state(s) which can accommodate a drug molecule. Prediction of such population shifts and non-native structures is non-trivial using current methods.$^{36}$

Recent work has utilized computational methods to identify regions of the protein that could harbor these cryptic allosteric sites using all-atom simulations.$^{37,38,39}$ Indeed, several possible binding sites were predicted on the antibiotic resistance protein TEM-1 β-lactamase using this method. These predicted sites may be attractive targets for drugs that inhibit this antibiotic resistance protein.$^{35,38}$ Experimentally validating these predicted sites, however, is essential. This thesis will discuss the use of thiol chemistry to label the high-energy states of TEM-1 β-lactamase in order to both validate these computationally predicted cryptic binding sites as well as further describe the conformations that give rise to such binding sites.

### 1.5 Folding Pathways and Intermediates

Of the multitude of conformations that are important to protein structure and stability, the conformations transiently populated during the folding process are critical. Indeed, for certain proteins it is thought that each molecule populates one or many such non-native structures as it folds to its native state. Population of such folding intermediates has been proposed to serve both as a nexus for aggregation and a productive step in folding to the native state.$^{40-44}$
While understanding the roles that these intermediates play remains a complicated problem, biophysical characterization has shown that both sequence and folding conditions can play a role in modulating the presence and stability of these folding intermediates.\textsuperscript{45-49} In certain cases folding pathways can be altered to remove intermediates through sequence variation or altered folding conditions.\textsuperscript{50-54} In addition, the structure of folding intermediates can be altered using small variations in sequence. The prototypical example of this phenomenon is apomyoglobin. This protein has a folding intermediate whose structured regions can be shifted using rationally designed point mutations despite the fact these variants still fold to the same native structure.\textsuperscript{55-59} A thorough characterization of the folding intermediates in many evolutionarily related proteins (Chapter 3) is important to disentangle the role that sequence, topology, and folding conditions play in determining the major features of the folding landscape.

1.6 Energy Landscapes of RNases H

Ribonuclease H (RNase H) from \textit{E. coli}, ecRNH, is a classic model system for studying the energy landscape. It is a 155-amino acid protein with mixed α/β structure.\textsuperscript{60} When studied by classic stopped flow methods (CD or fluorescence), folding of ecRNH\(^*\) (the * denotes a cysteine-free version of the protein) occurs in two phases. The first phase reaches completion during the 10 ms ‘dead-time’ of modern stopped flow instruments; a slower folding phase occurs on the order of seconds. The fast phase of folding has been shown to correspond to structure formation in the core region of the protein, Helices A-D and Strands 4 & 5 (herein referred to as I\(_{\text{core}}\) and illustrated in Figure 1-3).\textsuperscript{33,61} The rate-limiting step in folding of ecRNH\(^*\) involves the formation of structure in the remainder of the protein (Strands 1-3 and Helix E). This pathway is robust in many ecRNH\(^*\) variants despite the fact that a two-state folding version of the protein can be produced using a single point mutation.\textsuperscript{62,63} A diverse set of RNases H from many sources all populate an intermediate during folding, including those from HIV-1, \textit{Chlorobium tepidum} (ctRNH\(^*\)), and \textit{Thermus thermophilus} (ttRNH\(^*\)) suggesting a conserved folding pathway for this protein.\textsuperscript{47-49,64} In addition, several ancestral RNases H generated using ancestral sequence reconstruction on a phylogenetic tree linking ttRNH\(^*\) and ecRNH\(^*\) have been shown to fold through an intermediate, suggesting that the presence of an intermediate is conserved across evolutionary time.\textsuperscript{65} Recent work using a pulsed-labeling hydrogen exchange method showed that in ecRNH\(^*\), formation of the I\(_{\text{core}}\) intermediate prior to the rate limiting step does not take place in a single concerted step, but rather is the result of several regions of the protein consecutively gaining structure until the full I\(_{\text{core}}\) structure forms immediately prior to the rate-limiting folding step (Fig. 1-3).\textsuperscript{30} While the structures of the intermediate for several of the other RNase H variants have been studied, none have achieved the temporal or sequence resolution of this study. Hydrogen-exchange on related RNases H is, therefore, a promising technique with which to study the role of protein sequence in determining the structure of the folding intermediates of RNases H.
In addition to the kinetic folding intermediates, other aspects of the RNase H energy landscape have been studied extensively. The I\textsubscript{core} region of multiple RNases H have been shown to form a folded structure even in a truncated version where the remainder of the protein has been removed.\textsuperscript{66,67} For full-length ecRNH\textsuperscript{+}, a native-state hydrogen exchange study revealed that several partially unfolded forms (PUFs) of the protein exist in equilibrium with the native state under physiological conditions. Native-state hydrogen exchange probes the stability and structure of such PUFs by measuring both how protected each amide site is under native conditions and how this protection changes upon addition of small amounts of denaturant.\textsuperscript{26,27,68} For ecRNH\textsuperscript{+}, there are two such PUFs, or sub-global unfolding events - Strands 1, 2, 3, 5 and Helix E unfold at 7.4 kcal/mol, and Strand 4 and Helix B unfold at 8.7 kcal/mol. The remainder of the protein is labeled only when the protein samples the fully unfolded state with a stability of 10 kcal/mol relative to the native state.\textsuperscript{27} Similar studies revealed that both ttRNH\textsuperscript{+} and a thermostabilized ecRNH\textsuperscript{+} mutant, D10A, have a similar distribution of stability.\textsuperscript{69,70} Does the topology of the RNase H fold dictate this distribution of stability, or does sequence play a large role? Until now, addressing this question was difficult due to the laborious NMR detection involved in these hydrogen exchange studies. NMR detection is limited to proteins and variants that are amenable to NMR, have NMR backbone assignments and can be purified in large quantities. New techniques measuring hydrogen exchange with mass spectrometry open the possibility of probing many more proteins, thereby covering a much larger sequence space with the temporal and sequence resolution necessary to study these features of the energy landscape.

Figure 1-3. The folding pathway of ecRNH\textsuperscript{+}

The RNase H protein from E. coli is known to fold to its native structure in several sequential steps. A) The folding pathway of ecRNH\textsuperscript{+}. The first protected region of the protein is in Helix A/Strand 4 (blue). The next two steps in are protection Helix D/Strand 5 (green), and then the region of Helix B/C and the loop (yellow). Finally, the remainder of the protein (in red), adopts protection in the rate-limiting step for folding, giving the fully-folded native structure of the protein. B) Crystal structure (PDB: 2RN2) of ecRNH\textsuperscript{+} with coloration matching the individual folding steps.
1.7 Cotranslational Folding through the lens of HaloTag

Most studies on protein folding involve in vitro systems where the protein is refolded under dilute conditions in the test tube. Recent advances have begun to address how translation may affect this folding process. Translation occurs vectorially and much more slowly than the steps in protein folding. This suggests that every protein has the opportunity to sample possible native and non-native conformations while it is still being translated and tethered to the ribosome. Translation can affect the folding pathway of proteins,71–73 and these changes in folding pathway can alter the folding yield of protein.74,75 Studies on stalled ribosome-nascent chain complexes have shown that proximity to the ribosome can affect the stability of the nascent chain.76,77 What is lacking, however, is a detailed understanding of a protein’s co-translational folding pathway and how that pathway differs from the well-characterized refolding pathway.

An excellent model protein for studying this is the protein HaloTag, a modified haloalkane dehalogenase originally designed to allow specific protein targeting of a wide variety of ligands. Under standard conditions, binding of a ligand to HaloTag is very fast.78 Since both folding and translation occur much more slowly, this binding can serve as a probe for a fully-folded HaloTag even in the complex mixture of molecules present during translation. Use of a fluorescently-labeled ligand allows measurement of the folding of HaloTag both co-translationally and in dilute solution, and labeling of high-energy states can illustrate how the energy landscape of a protein can be modulated during translation by the ribosome.79

1.8 Mass Spectrometry and Data Analysis

As discussed above, exchange of amide hydrogens in a protein can serve as a useful probe for protein structure and stability. Detecting hydrogen exchange with mass spectrometry (HX-MS) has several advantages over measurement with NMR. The freedom to perform an experiment without the need to elucidate NMR assignments, combined with the small quantity of protein required for mass spectrometry allow experiments to be quickly performed on many unique proteins. The decreased work-up time between labeling and detection allows more probes to be identified and monitored. This was illustrated in the case of the pulsed-labeling study on ecRNH*, where previous NMR work could not monitor folding in the loop region of the protein, but the MS experiment was able to unambiguously show that the loop achieves native protection prior to the rate-limiting step in folding.30,33 In addition, since mass spectrometry requires detection at the peptide level; this preserves information about cooperative exchange events that are lost in the NMR data as each site is measured individually.

While these advantages make HX-MS a very powerful technique, they also come with caveats. The novelty of mass spectrometry detection in HX experiments means that data analysis methods focused on MS data are still incomplete and lag far behind those developed for NMR. This is further complicated by inherent differences in measurement by MS and NMR. Whereas NMR data collection allows, and indeed requires, detection of exchange at the scale of a single residue, mass spectrometry detects several residues together, and it can be difficult to determine whether these residues exchange with a single rate or several similar rates. In limiting cases where peptides unambiguously exchange with two or more rates, it is difficult to determine which residue is exchanging with which rate, making quantitative treatment of protection
much more difficult than with NMR. Using mass spectrometry to assign protection factors similar to NMR experiments will thus require novel methods of interpreting MS data.

Previous efforts have achieved near-residue resolution by identifying which residues contain deuterium at the sub-peptide level using overlapping peptides. These methods require hundreds of peptides redundantly covering the protein dozens of times over, but such coverage is still the exception in the field. Additional work to characterize protection using MS detection has shown that the average amount of protection across an individual peptide can be quantified empirically without the need for identifying individual exchange events. However, the average protection across a peptide can be less useful than site-resolved protection since large peptides can cover several regions of a protein with vastly different protection factors. These limitations represent a bottleneck in turning the massive datasets produced by HX-MS into quantitative characterizations of protein structure, and expanding the existing methods to obtain protection factors with better resolution and in more general cases is necessary for a quantitative treatment of HX-MS data.

1.9 Summary

The work presented in this thesis uses thiol and hydrogen exchange chemistries to label sites that are protected in the native state in order to identify and characterize non-native states of several proteins. In Chapter 2, experiments using thiol labeling to identify possible cryptic binding sites in the protein β-lactamase are discussed. These sites agreed well with known and computationally predicted binding sites. This work was published in PNAS with me as the second author (I carried out the experimental studies). In Chapter 3, the use of HX-MS to study the folding intermediates in several variants of RNase H to elucidate the role of sequence in determining the structure of these transient states is discussed. The results of these experiments showed that while the presence of one intermediate, Icore, is preserved across billions of years of evolutionary divergence, the pathway leading to this structure was allowed to vary. Further work showed that a local property of the sequence, its intrinsic helicity, is capable of modulating the earliest folding events. This work is currently in review at eLife and I am co-first author. In Chapter 4, a similar experiment was employed to identify a folding intermediate in the protein HaloTag that may explain the difference in its folding pathway between cotranslational folding and refolding in dilute solution. This work was published in Science Advances. I carried out the pulsed labeling hydrogen exchange and am second author. Finally, in Chapter 5, data analysis methods that can improve the quantitation of protection in HX-MS experiments are discussed. This work will be submitted for publication and I am the first author.
1.10 References


30. Hu, W. et al. Stepwise protein folding at near amino acid resolution by hydrogen


74. Frydman, J., Erdjument-Bromage, H., Tempst, P. & Hartl, F. U. Co-translational domain folding as the structural basis for the rapid *de novo* folding of firefly


Chapter 2:
Identifying cryptic binding sites: combining excited-state labeling with computation to reveal novel drug binding sites
This article was published in PNAS in 2015, and I am second author on this work. My contribution to this work consists of the experiments, analysis, and interpretation of work on cysteine labeling, unfolding, and activity measurements of several of the model proteins including labeling within the first predicted binding pocket as well as several controls (M182T, L286C, L190C, I260C, and A150C).

2.1 Abstract

The discovery of drug-like molecules that bind pockets in proteins that are not present in crystallographic structures yet exert allosteric control over activity has generated great interest in designing pharmaceuticals that exploit allosteric effects. However, there have only been a small number of successes, so the therapeutic potential of these pockets—called hidden allosteric sites—remains unclear. One challenge for assessing their utility is that rational drug design approaches require foreknowledge of the target site, but most hidden allosteric sites are only discovered when a small molecule is found to stabilize them. We present a means of decoupling the identification of hidden allosteric sites from the discovery of drugs that bind them by drawing on new developments in Markov state modeling that provide unprecedented access to microsecond- to millisecond-timescale fluctuations of a protein’s structure. Visualizing these fluctuations allows us to identify potential hidden allosteric sites, which we then test via thiol labeling experiments. Application of these methods reveals multiple hidden allosteric sites in an important antibiotic target—TEM-1 β-lactamase. This result supports the hypothesis that there are many as yet undiscovered hidden allosteric sites and suggests our methodology can identify such sites, providing a starting point for future drug design efforts. More generally, our results demonstrate the power of using Markov state models to guide experiments.

2.2 Introduction

A hidden allosteric site is a binding pocket that is not present in the crystal structure of a protein, but becomes available as the protein fluctuates and is capable of controlling the protein’s function by communicating with the active site (Fig. 2-1). Ligands that bind these sites exert control over the protein’s function by perturbing the ensemble of structures the protein adopts. Such sites could have unknown biological functions and serve as valuable targets for drug design, particularly for proteins that are currently considered undruggable because known structures lack pockets that are suitable for drug design. Unfortunately, it has been difficult to explore either of these possibilities because identifying hidden allosteric sites and molecules that bind them remains a profound challenge. For example, most structure-function studies focus on a single representative structure of a protein and give little hint as to where hidden allosteric sites might occur or what sort of molecules might bind them.

Detecting hidden allosteric sites experimentally is difficult because most of the available methods couple the identification of such sites with the drug discovery process. Given the difficulties inherent to drug design, this coupling likely produces many false negatives, leaving hidden allosteric sites undiscovered. For example, high-
throughput screening can reveal hidden allosteric sites\textsuperscript{5,6} but failure to identify allosteric modulators does not disprove the existence of allosteric sites. Tethering provides a site-directed screen that is useful for specifically searching for allosteric sites but will still suffer from false negatives if the library being screened does not include small molecules that will bind an allosteric site tightly enough.\textsuperscript{7,8} Understanding the full ensemble of structures a protein can adopt would overcome these limitations,\textsuperscript{9} but such an understanding remains elusive. For instance, room temperature crystallography and NMR have the potential to reveal alternative structures containing hidden allosteric sites, but further developments are required to make such measurements routine for any given protein target.\textsuperscript{10–15}

A number of computational techniques have been developed to aid in the discovery of hidden allosteric sites. For example, there are a variety of methods for understanding how information flows from one region of a protein to another,\textsuperscript{16–20} as well as for identifying potential binding pockets.\textsuperscript{21–25} More recent work has accounted for both of these ingredients.\textsuperscript{26,27} Although these methods are important developments, many are only applicable to small proteins operating on fast timescales. Therefore, new approaches are needed for addressing many biologically relevant systems. Furthermore, these methods only partially decouple the discovery of hidden allosteric sites from the identification of allosteric ligands because identifying a compound that binds to a predicted site is still the primary means of testing computationally predicted allosteric sites.

Here, we use a combination of computation and experiment to identify hidden allosteric sites without requiring the simultaneous discovery of ligands that bind and modulate them. Decoupling the discovery of allosteric sites from the identification of

\textbf{Figure 2-1. Crystal structures of TEM-1 β-lactamase}

The crystal structures of TEM-1 β-lactamase in the absence of any ligand (blue) and with an inhibitor (cyan) bound in a hidden allosteric site (yellow). A key catalytic serine (S70) is in green spheres.
allosteric ligands should facilitate drug discovery by providing more information to base design decisions on. For example, rather than performing blind screens, computationally generated structures of potential allosteric sites can be used as a starting point for rational design. The locations of potential hidden allosteric sites can also be used to direct tethering screens, potentially providing a less resource-intensive means of discovering hidden allosteric sites than applying tethering to every possible location in a protein. Our approach can be applied to most soluble proteins, but could be of particular value for the large number of proteins that are currently considered undruggable because their active sites or binding interfaces are not considered viable drug targets.

The first step in our approach is to build a Markov state model of a protein of interest, which is essentially an atomically detailed map of the ensemble of conformations the protein can adopt. Such models are of great value because they can capture relatively slow conformational changes that are typically beyond other computational methods and they allow the user to learn what degrees of freedom are important rather than requiring the user to select them a priori. Potential allosteric sites are then identified by querying the Markov model for local fluctuations that form pockets that are surrounded by residues whose rotameric orientations are correlated with those of the active site. These correlations are a property of the ensemble of structures the protein can adopt and give insight into where perturbations, such as ligand binding, are likely to exert allosteric control over distant sites, such as the active site.

Predicted pockets are tested experimentally with thiol labeling. In thiol labeling, a cysteine is introduced at a site buried within a pocket and a chemical reagent is introduced, such as 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, also called Ellman's reagent). This reagent will react covalently with the cysteine thiol if the thiol is sufficiently exposed and available to solvent. DTNB is also attractive for our purposes because it is a drug-sized molecule, so its ability to bind within a pocket suggests there may be enough room for a typical drug to bind. Thiol labeling is analogous to hydrogen-deuterium exchange but is sensitive to changes in the exposure of side-chains rather than the backbone. Therefore, it is better suited to detecting pockets that form when elements of secondary structure separate from one another, as is observed in many pocket-opening fluctuations, without requiring unfolding and exposure of the backbone amides. This approach closely parallels thiol-exchange techniques for studying protein folding/unfolding and can provide opening rates for the various pockets as well as the fraction of time the pocket is open. Communication between a pocket and the active site is tested by assaying whether covalently linking a small molecule within a pocket has any detectable effect on the protein's activity, as in tethering. A number of controls are also performed to ensure that the labeling we observe is due to local fluctuations captured by our computational models rather than global unfolding.

To test this approach, we apply it to TEM-1 β-lactamase with the M182T substitution, a change commonly found in antibiotic-resistant variants that is known to stabilize the native state. This protein is an important drug target because of its role in
antibiotic resistance. It already has one known hidden allosteric site, so it is an excellent system for testing whether our approach can discriminate allosteric sites from nonallosteric sites. Finally, β-lactamase is a large enough protein that it is reasonable to ask if there are other hidden allosteric sites, especially given our recent computational prediction that such sites exist. Newly predicted allosteric sites that are similar in nature to the known site could be attractive drug targets given that we already know ligands that bind the known hidden allosteric site are capable of modulating β-lactamase’s activity.

2.3 Results and Discussion

**Computation and Experiment Successfully Detect the Known Hidden Allosteric Site**

Our previous work demonstrated that our computational approach successfully identifies the known hidden allosteric site in TEM-1 β-lactamase. The known site is the largest of the unexpected pockets formed in this protein and is also in the open state more than any other pocket. The conformations of residues surrounding this pocket are also correlated with the conformations of active site residues, providing a means for communication between these sites. Therefore, we expect perturbations at this site can modulate the ensemble of structures β-lactamase adopts in a manner that alters the protein’s activity. These results suggest an important role for conformational selection in the function of this hidden allosteric site. That is, the pocket is present even in the

![Figure 2-2 Thiol labeling of the known hidden allosteric site](image)

(A and B) Surface representations of the closed and open states of the known hidden allosteric site, respectively. L286 (yellow) is only visible in the open state. A key catalytic serine (S70) is shown in green as a reference point.

(C) Labeling of L286C in 1 mM DTNB. (D) The dependence of the labeling rate of L286C on the concentration of the labeling reagent (DTNB) with error bars representing the SD from three replicates.
absence of a small molecule to bind it rather than being created by interactions between the protein and small molecule, which is often called induced fit. Therefore, the known hidden allosteric site should be detectable in our thiol labeling experiments because the pocket will be present even in the absence of a specific allosteric modulator.

If the known hidden allosteric site opens as a result of the fluctuations predicted by our computational model, then we expect to see labeling of residues whose side-chains line this pocket. To test this prediction, we applied our thiol labeling technique to L286. This residue was chosen because it has negligible solvent accessible surface area in the apo structure of β-lactamase, but our computational model predicts that it becomes exposed when the known hidden allosteric site opens (Fig. 2-2 A and B). The L286C mutation required for thiol labeling is also one of the more conservative mutations we could have chosen in the known hidden allosteric site, so this mutation minimizes potential perturbation to the protein.

Indeed, our labeling studies indicate that the known hidden allosteric site opens as β-lactamase fluctuates. The L286C variant labels at a rate of $\sim 5.9 \times 10^{-4} \pm 5.7 \times 10^{-5}$ s$^{-1}$ in 1 mM DTNB (Fig. 2-2C and Fig. 2S-1). The expected rate of labeling for a fully exposed residue is about 1 s$^{-1}$, so the observed labeling cannot be attributed to a reorganization of the protein’s structure that exposes this residue in the ground state.

As in hydrogen exchange, we interpret the observed labeling rate with the Linderstrom–Lang model. This model assumes the protein is in equilibrium between conformations where a pocket is either closed or is open and available to react with DTNB (Scheme 2-1). Given an opening rate ($k_{\text{op}}$), closing rate ($k_{\text{cl}}$), and rate of labeling from the open state ($k_{\text{int}}$), the observed rate of labeling is $k_{\text{obs}} = (k_{\text{op}}k_{\text{int}})/(k_{\text{op}} + k_{\text{cl}} + k_{\text{int}})$. In the limit where $k_{\text{op}} \ll k_{\text{cl}}$ and $k_{\text{cl}} \ll k_{\text{int}}$, this reduces to $k_{\text{obs}} = k_{\text{op}}$. This scenario is called the EX1 regime and can be identified because the observed rate of labeling will be independent of the concentration of labeling reagent. In the limit where $k_{\text{op}} \ll k_{\text{cl}}$ and $k_{\text{cl}} \gg k_{\text{int}}$, then $k_{\text{obs}} = K_{\text{op}}k_{\text{int}}$, where $K_{\text{op}} = k_{\text{op}}/k_{\text{cl}}$ is the equilibrium constant for the pocket being open. This scenario is called the EX2 regime and can be identified because the observed rate of labeling will be linearly dependent on the concentration of labeling reagent ($k_{\text{int}}$).

$$
\begin{align*}
\text{Closed} & \overset{k_{\text{op}}}{\longrightarrow} \text{Open} & \overset{k_{\text{int}}}{\longrightarrow} \text{Labeled} \\
\text{Open} & \overset{k_{\text{cl}}}{\longrightarrow} \text{Closed}
\end{align*}
$$

Scheme 1

To determine whether the observed rate of labeling is providing information about the opening rate or the fraction of time a pocket is open or exposed, we measured the rate of labeling with varying concentrations of DTNB. Figure 2-2D shows that the labeling rate is independent of [DTNB], which is consistent with the EX1 regime. Therefore, we conclude that the observed rate of labeling captures the opening rate of this pocket.
Pockets are Clearly Distinguishable from Nonpockets

Our experimental approach might give false positives if the cysteine mutations cause significant destabilization of the protein. For example, introducing a cysteine could globally destabilize the protein such that labeling occurs directly from global unfolding rather than transient exposure of the pocket within the native state ensemble. If this was true for the L286C variant, we would expect the labeling rate to approximate the rate of global unfolding because labeling is in the EX1 regime.

To test whether labeling is due to global unfolding, we determined the unfolding rate of our cysteine variant and compared it with the measured labeling rate. Following previous work on the unfolding of β-lactamase, we measured the unfolding rate of the L286C variant by monitoring the change in the circular dichroism (CD) signal as a function of the final urea concentration (Fig. 2-3). Extrapolating back to 0 M urea (the labeling conditions), we find that the rate of unfolding is about 20-fold smaller than the observed rate of labeling. Therefore, labeling must be due to a fluctuation across a barrier from the native state that is lower than the barrier to global unfolding.

As a control, we created cysteine variants at buried sites not predicted to form a pocket. Residues L190 and I260 are both buried in the ligand-free structure of β-lactamase, and our model predicts that there are no pockets that expose these residues to drug-sized molecules. Consistent with this prediction, we do not observe any labeling of cysteines at these positions over the course of a 12-h labeling reaction. Therefore, we conclude that these residues remain buried in the native-state ensemble and that introducing a cysteine does not cause a local destabilization that creates an unpredicted pocket or local unfolding. This result, in combination with the lack of observed labeling for the two endogenous cystines in the protein that are oxidized in a disulfide bond, also confirms that the labeling we observe for other residues is not due to a reaction with the two cysteines that naturally form a disulfide in β-lactamase. The fact that our computational model successfully discriminates where labeling will and will not occur also adds significant weight to our conclusion that labeling is due to the formation of a pocket rather than a large-scale unfolding event.

Given the proximity of the known hidden allosteric site to two of the four tryptophan residues in β-lactamase, we reasoned that opening of this pocket may expose these trytophans to solvent and lead to a change in the protein’s fluorescence. Indeed, opening of this pocket in our computational model increases the solvent accessible surface area of Trp229’s side-chain from 36% in the ligand-free structure to 69 ± 9% when the pocket is open. The solvent accessible surface area of Trp290’s side-chain increases from 43% in the ligand-free structure to 85 ± 8% when the pocket is open. Because pocket opening precedes global unfolding and might be on the pathway to global unfolding, we hypothesized that monitoring unfolding by fluorescence should detect pocket opening and yield a faster rate than monitoring unfolding by CD. To test this prediction experimentally, we measured the rate of change in fluorescence of the L286C variant as a function of the final urea concentration and used linear extrapolation to find the rate of change in the absence of denaturant. This procedure yielded a rate of $5.5 \times 10^{-4} \pm 2.4 \times 10^{-4} \text{s}^{-1}$, in reasonable agreement with the rate of labeling with DTNB.
of $5.9 \times 10^{-4} \pm 5.7 \times 10^{-5}$ s$^{-1}$. The fact that these rates are 20-fold larger than the rate of unfolding demonstrates that labeling precedes unfolding and, therefore, occurs from a rare state on the native side of the rate-limiting barrier to unfolding. Interestingly, this state appears to be distinct from previously characterized high-energy states on the unfolded side of the rate-limiting barrier to unfolding that were also detected by fluorescence.\textsuperscript{38,39}

Figure 2-3. Labeling and Unfolding of L286C
Thiol labeling is not due to unfolding. Log of the unfolding rate of L286C as monitored by CD for different urea concentrations with a linear fit (black line) used for extrapolating back to the unfolding rate at 0 M urea. The labeling rate (yellow circle) is considerably faster than unfolding, so it must correspond to a fluctuation within the native state.
There is communication between the known allosteric site and active site

We also exploited our thiol labeling to test whether there is communication between the pockets we detect and the active site, as indicated by correlations in the ensemble of structures β-lactamase adopts. We previously predicted that almost any hidden binding pocket should also serve as an allosteric site due to coupling of many residues to different portions of the active site.27 To test whether a given pocket communicates with the active site, we measured the activity of proteins with and without TNB (one half of DTNB) covalently bound within the pocket. A measurable change in activity would demonstrate that there is communication, although it should not be used as a quantitative measure of potential inhibition because other molecules that bind the same site could be more potent inhibitors or even enhance the protein’s activity.40 Using this approach, we find that the specific activity of L286C is reduced from 361 ± 29 to 97 ± 6 nmol product/μg/min. These results suggest that there is communication between the site of modification and the active site. Although DTNB is a drug-sized molecule (Materials and Methods), TNB is significantly smaller than typical drugs and has not been optimized for binding this hidden allosteric site. Therefore, it is entirely possible that an allosteric modulator specifically designed to bind this site could be a much stronger β-lactamase inhibitor. Identifying such noncovalent inhibitors (or activators) would serve as the ultimate verification of the existence of our hidden allosteric sites and remains an important future direction. Although we have not yet discovered new molecules that bind the hidden allosteric sites revealed by our approach, we note that the allosteric inhibitor discovered by Horn et al. demonstrates that it is possible for small molecules to bind such hidden allosteric sites strongly enough to stabilize the open form of a pocket and alter an enzyme’s activity.35

<table>
<thead>
<tr>
<th>Pocket</th>
<th>Mutation</th>
<th>Unlabeled Activity</th>
<th>Labeled Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Background (M182T)</td>
<td>532 ± 5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>L286C</td>
<td>361 ± 29</td>
<td>97 ± 6</td>
</tr>
<tr>
<td>2</td>
<td>A232C</td>
<td>545 ± 75</td>
<td>255 ± 21</td>
</tr>
<tr>
<td>2</td>
<td>A249C</td>
<td>193 ± 13</td>
<td>150 ± 18</td>
</tr>
<tr>
<td>3</td>
<td>S203C</td>
<td>520 ± 13</td>
<td>345 ± 17</td>
</tr>
<tr>
<td>Surface</td>
<td>A150C</td>
<td>235 ± 25</td>
<td>249 ± 28</td>
</tr>
<tr>
<td>Buried</td>
<td>L190C</td>
<td>213 ± 22</td>
<td></td>
</tr>
<tr>
<td>Buried</td>
<td>I260C</td>
<td>267 ± 38</td>
<td></td>
</tr>
</tbody>
</table>

Table 2-1 Specific activities (nmol product/μg protein/min) of labeled and unlabeled proteins

As a control, we tested whether thiol labeling of residues that our computational model predicts should have little communication with the active site alters β-lactamase’s activity. Specifically, we chose to perform thiol labeling of A150 because it is a surface residue that is available for labeling and surrounding residues have weak correlations with the active site in our computational model. Complete labeling of the A150C variant
has a negligible effect on the protein’s activity (Table 2-1), consistent with our prediction that it does not communicate with the active site.

**Discovery of Previously Unidentified Hidden Allosteric Sites**

Now we can begin testing whether our model successfully predicts novel hidden allosteric sites. Toward this end, we chose to focus on pockets that our simulations predict will expose residues that are completely buried in the static, apo-crystal structure. Such pockets are the most amenable to our thiol labeling experiments, although there may be other pockets that are equally potent allosteric sites but lack residues with this differential exposure.

Residues appropriate for our thiol labeling experiments were selected by examining the fraction of each residue’s surface area that is typically accessible to probes of varying sizes (Fig. 2-4). Specifically, we chose probe radii of 0.14, 0.24, 0.34, 0.44, and 0.54 nm. This range mimics molecules like water at the smallest scale and more drug-like molecules at the largest scale. We calculated the accessible surface area for the side-chain of each residue for a given probe size as follows: (i) computationally roll a sphere with the given probe radius across the surface of a representative structure for each state in the Markov model and calculate the accessible surface area of every residue’s side-chain; (ii) calculate the average accessible surface area of each side-chain by taking the average across all states, weighted by their equilibrium populations; and (iii) divide the result for each residue by the total possible accessible surface area of its side-chain. One result of this analysis is that the fluctuations β-lactamase undergoes make basically every residue’s side-chain accessible to water, as previously observed in other proteins. However, many residues are not accessible to larger molecules. We selected residues that are accessible to probes with radii of at least 0.34 nm because this is consistent with the size of DTNB and residues that we have already shown to label are accessible at this probe size, whereas residues that we have shown do not label are not. Of the residues that meet this criterion, we selected the smallest residues possible to minimize the

![Figure 2-4. Residues selected for labeling in each pocket](image)

(A) Ribbon diagram of β-lactamase highlighting residues in the known hidden allosteric site (L286, yellow), the first predicted site (A232, red), and the second predicted site (S203, magenta). A key catalytic serine (S70) is shown in green. (B) Average percent of residues’ surface area that is accessible to a variety of probe sizes. L190 and I260 are buried, whereas A150 is on the surface.

27
perturbation caused by mutating to a cysteine. Based on these criteria, residues A232 and A249 point into the most promising pocket. Residue S203 also points into a second pocket.

We performed thiol labeling of A232C to test our first predicted pocket (Fig. 2-5). Figure 2-5 C and D show that the A232C variant labels at a rate of $3.6 \times 10^{-3} \pm 1.0 \times 10^{-3} \text{ s}^{-1}$ independent of the concentration of labeling reagent, and therefore this is the rate at which the residue becomes exposed. The rate of unfolding is also 200-fold slower than the rate of labeling (Table 2-2), so the observed labeling is not due to global unfolding. Complete labeling of the protein reduces the activity of the protein by 1.5-fold (Table 2-1). In addition to our labeling experiments, we again reasoned that opening of this pocket could lead to a change in fluorescence by exposing Trp229 to solvent. Indeed, opening of this pocket in our computational model increases the solvent accessible surface area of Trp229’s side-chain from 36% in the ligand-free structure to $56 \pm 12\%$ when the pocket is open. Experimentally monitoring unfolding by fluorescence, as described previously, yielded a rate of $1.9 \times 10^{-3} \pm 1.2 \times 10^{-3} \text{ s}^{-1}$, in reasonable agreement with the DTNB labeling rate for this variant. We also performed separate experiments on an A249C variant. The side-chain of this residue points into the same pocket but has less exposure than residue 232 because it is not exposed in the exact same set of structural states where this pocket is open as residue 232. Indeed, we observe labeling of the A249C variant at a rate fivefold less than the A232 variant (Table 2-S1), consistent with the residue at position 249 being exposed on
opening of the pocket. Based on all of the results for these two variants, we conclude that this site is a hidden allosteric site, consisting of an unexpected pocket with the ability to communicate with the active site. An allosteric modulator of this site would need to have a greater effect on activity but, as explained before, this is entirely possible.

We also tested the second predicted pocket via thiol labeling of S203C. This residue labels at a rate of $1.5 \times 10^{-2} \pm 3.4 \times 10^{-3}$ s$^{-1}$, again independent of the concentration of labeling reagent (Fig. 2-S2). This rate is 50-fold faster than the rate of global unfolding (Table 2-S1), so it captures the rate of exposure of the residue. There is also communication between this site and the active site, as demonstrated by an $\sim 1.5$-fold reduction in activity on labeling. Therefore, we conclude that this second predicted site is also a hidden allosteric site.

### 2.4 Conclusions

We developed an approach that combines computation and experiment to detect hidden allosteric sites arising from the ensemble of structures a protein can adopt. Importantly, our approach does not require the simultaneous discovery of small molecules that bind and modulate these sites, so our methodology can be used to guide subsequent drug design efforts.

Using this approach, we have demonstrated that a single protein—TEM-1 β-lactamase—accommodates multiple hidden allosteric sites. This result is surprising because TEM-1 β-lactamase has been studied extensively without observing these sites. Furthermore, there may even be other hidden allosteric sites in this single protein that are not amenable to the experimental methodology we describe here.

Our results suggest there are many as yet undiscovered hidden allosteric sites and that our techniques should provide a means of detecting them. Once discovered, these allosteric sites can then be targeted with rational drug design or followed up on to discover their biological relevance. In the case of TEM-1, the hidden allosteric sites we discovered could be valuable targets for antibiotic development.

These results lay an important foundation for future work on hidden allosteric sites. For example, an important next step will be to discover new allosteric modulators that bind the hidden allosteric sites revealed by our methodology. Furthermore, our results demonstrate the value of our advanced computational methods and argue for further developments to make an even more quantitative comparison between computation and experiment.

### 2.5 Materials and Methods

#### Pocket Prediction.

As described previously,27 1,000 simulations were run with Gromacs$^{42,43}$ deployed on the Folding@home distributed computing platform$^{44}$ starting from Protein Data Bank ID code 1JWP,$^{45}$ for an aggregate of 81 μs of dynamics. Each simulation was run at 300 K using the Amber03 force field with explicit TIP3P water and seven
sodium ions to neutralize the charge.\textsuperscript{46} V-sites were used to allow for a 5-fs time step. The Markov state model was constructed with MSMBuilder.\textsuperscript{47,48} Following a standard protocol,\textsuperscript{49} every 10th conformation from the simulations for each protein were clustered with a k-centers algorithm based on the RMSD between Cα and Cβ atoms until every cluster had a radius—i.e., maximum distance between any data point in the cluster and the cluster center—less than 1.2 Å. Then, 10 sweeps of a k-medoids update step were used to center the clusters on the densest regions of conformational space. The remaining 90\% of the data were then assigned to these clusters, and states with only inbound or outbound transitions were discarded. Based on the model’s implied timescales, a lag time of 2 ns was used. The fraction of time a residue is exposed for a given probe radius was calculated as the sum of the equilibrium populations of all of the states the residue is exposed in. Accessible surface areas were measured with Gromacs and a Voronoi-based method.\textsuperscript{50} Structures were visualized with PyMOL.\textsuperscript{51}

**Protein Expression and Purification.**

TEM-1 β-lactamase with the M182T stabilizing mutation and all cysteine variants of this background sequence were purified from the periplasmic fraction of BL21(DE3) cells as described previously.\textsuperscript{52} Cysteine mutations were introduced with Quik-Change mutagenesis. BL21(DE3) Escherichia coli transformed with β-lactamase mutant plasmids were grown at 37 °C to OD\textsubscript{600} = 0.6 and then cooled to 18 °C for 30 min before a 16-h induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were pelleted and resuspended in 50 mL/L of culture of 30 mM Tris, 20\% (wt/vol) sucrose, pH = 8, and stirred for 10 min at room temperature. After centrifugation (20 min at 10,000 × g, 4 °C), the pellet was resuspended in an equal volume of ice-cold 5 mM MgSO\textsubscript{4} and stirred for 10 min at 4 C. Centrifugation (20 min at 10,000 × g, 4 °C) produces a supernatant that contains the periplasmic fraction of the cells. This supernatant is dialyzed into 2 × 4 L of 20 mM sodium acetate, pH = 5.5, and 0.2 μM filtered. After filtration, the protein is bound to a HiTrap CaptoS column and eluted with a 50-column volume (CV) gradient from 0 to 0.6 M NaCl. Fractions that contain β-lactamase activity are then pooled, concentrated, and run over a Hi-Load 16/60 Superdex 75 column in 20 mM Tris, pH = 8.0. Protein was stored at 4 °C in 20 mM Tris, pH 8.0.

**Activity Measurements.**

We measured enzyme activities following previous work.\textsuperscript{53} Protein activity was calculated using nitrocefin as a substrate [50 mM potassium phosphate, 10\% (vol/vol) glycerol, pH = 7.0]. Protein at a concentration of ~3 nM was equilibrated at 25 °C, and then nitrocefin was added from a 5 mg/mL stock in DMSO to a final concentration of 0.05 mg/mL and manually mixed. The reaction progress was monitored at 482 nm, and activity was calculated using an extinction coefficient at 482 nm of 15,000 M\textsuperscript{−1}·cm\textsuperscript{−1}.

**DTNB Labeling.**
Thiol labeling experiments were run with 30 μM protein and varying concentrations of DTNB (also called Ellman’s reagent) in 20 mM Tris, pH 8.0. The time course for labeling was monitored by following the absorbance at 412 nm in a Cary 100Bio UV-Vis spectrophotometer (27 °C) after starting the reaction via manual mixing. Rates were interpreted using the Linderstrom–Lang model.\textsuperscript{37}

DTNB has a molecular mass of 396.4 Da. Given that Lipinski’s rules, which are commonly used to define druglike molecules, include having a molecular mass less than 500 Da, DTNB is a drug-sized molecule. Furthermore, thiols react with DTNB via nucleophilic attack of DTNB’s disulfide, so the whole DTNB molecule must gain access to the pocket to react with one of our introduced cysteines. Therefore, observing labeling of a cysteine implies that the pocket that exposes that residue is large enough to accommodate a drug-sized molecule. After the reaction occurs, however, only a single TNB (half of a DTNB) will remain covalently bound to the protein. Therefore, TNB binding may have little effect on a protein’s structure and activity compared with drug-sized molecules that may bind to the same site.

To determine whether the labeling of a cysteine is in the EX1 or EX2 regime, the concentration of DTNB was varied by at least threefold, as has been done in previous studies (10–12). Using significantly lower DTNB concentrations was not possible while maintaining an excess of DTNB over protein. Using significantly higher DTNB concentrations was not possible because doing so would saturate the detector, making it impossible to observe labeling.

The bimolecular rate constant for labeling an exposed thiol at pH 8.5 is 7,000–8,000 M\textsuperscript{-1} s\textsuperscript{-1} and drops by about 10-fold for every unit decrease of the pH (10, 13). Therefore, the intrinsic rate of labeling for 1 mM DTNB at pH 8 is about 1 s\textsuperscript{-1}.

Equilibrium Melts and Unfolding Rates.

CD data were collected using an Aviv 410 spectropolarimeter with a Peltier temperature-controlled sample holder and an Aviv 62A DS spectropolarimeter with an 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 [Urea] at 27 °C. Samples were allowed to equilibrate overnight, and the signal for each sample was averaged over 60 s. For unfolding rates, 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.

Fluorescence data were acquired on a Jvon Horiba Fluoromax-3 with excitation at 280 nm (1-nm slit width) and emission at 340 nm (2-nm slit width) with an integration time of 0.1 s at T = 27 °C.

Rates were measured on dilution to final urea concentrations from 4 to 6 M urea. The rate in the absence of urea was determined by fitting the log of the rates as a function of urea to a line and extrapolating back to 0 M urea. Error bars were obtained from bootstrapping.
2.6 References


2.7 Supplemental Materials

Materials and Methods

Simulations and Modeling. As described previously, 1,000 simulations were run with Gromacs 2,3 deployed on the Folding@home distributed computing platform 4 starting from Protein Data Bank ID code 1JWP, 5 for an aggregate of 81 μs of dynamics. Each simulation was run at 300 K using the Amber03 force field 6 with explicit TIP3P water and seven sodium ions to neutralize the charge. V-sites were used to allow for a 5-fs time step.

The Markov state model was constructed with MSMBuilder 7,8 Following a standard protocol, 9 every 10th conformation from the simulations for each protein were clustered with a k-centers algorithm based on the RMSD between Cα and Cβ atoms until every cluster had a radius—i.e., maximum distance between any data point in the cluster and the cluster center—less than 1.2 Å. Then, 10 sweeps of a k-medoids update step were used to center the clusters on the densest regions of conformational space. The remaining 90% of the data were then assigned to these clusters, and states with only inbound or outbound transitions were discarded. Based on the model’s implied timescales, a lag time of 2 ns was used. The fraction of time a residue is exposed for a given probe radius was calculated as the sum of the equilibrium populations of all of the states the residue is exposed in.

Protein Expression and Purification. BL21(DE3) Escherichia coli transformed with β-lactamase mutant plasmids were grown at 37 °C to OD600 = 0.6 and then cooled to 18 °C for 30 min before a 16-h induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were pelleted and resuspended in 50 mL/L of
culture of 30 mM Tris, 20% (wt/vol) sucrose, pH = 8, and stirred for 10 min at room temperature. After centrifugation (20 min at 10,000 × g, 4 °C), the pellet was resuspended in an equal volume of ice-cold 5 mM MgSO$_4$ and stirred for 10 min at 4 °C. Centrifugation (20 min at 10,000 × g, 4 °C) produces a supernatant that contains the periplasmic fraction of the cells. This supernatant is dialyzed into 2 × 4 L of 20 mM sodium acetate, pH = 5.5, and 0.2 μM filtered. After filtration, the protein is bound to a HiTrap CaptoS column and eluted with a 50-column volume (CV) gradient from 0 to 0.6 M NaCl. Fractions that contain β-lactamase activity are then pooled, concentrated, and run over a Hi-Load 16/60 Superdex 75 column in 20 mM Tris, pH = 8.0. Protein was stored at 4 °C in 20 mM Tris, pH 8.0.

Activity Measurements. Protein activity was calculated using nitrocefin as a substrate [50 mM potassium phosphate, 10% (vol/vol) glycerol, pH = 7.0]. Protein at a concentration of ~3 nM was equilibrated at 25 °C, and then nitrocefin was added from a 5 mg/mL stock in DMSO to a final concentration of 0.05 mg/mL and manually mixed. The reaction progress was monitored at 482 nm, and activity was calculated using an extinction coefficient at 482 nm of 15,000 M$^{-1}$·cm$^{-1}$.

DTNB Labeling. DTNB has a molecular mass of 396.4 Da. Given that Lipinski’s rules, which are commonly used to define druglike molecules, include having a molecular mass less than 500 Da, DTNB is a drug-sized molecule. Furthermore, thiols react with DTNB via nucleophilic attack of DTNB’s disulfide, so the whole DTNB molecule must gain access to the pocket to react with one of our introduced cysteines. Therefore, observing labeling of a cysteine implies that the pocket that exposes that residue is large enough to accommodate a drug-sized molecule. After the reaction occurs, however, only a single TNB (half of a DTNB) will remain covalently bound to the protein. Therefore, TNB binding may have little effect on a protein’s structure and activity compared with drug-sized molecules that may bind to the same site.

To determine whether the labeling of a cysteine is in the EX1 or EX2 regime, the concentration of DTNB was varied by at least threefold, as has been done in previous studies.\(^{10-12}\) Using significantly lower DTNB concentrations was not possible while maintaining an excess of DTNB over protein. Using significantly higher DTNB concentrations was not possible because doing so would saturate the detector, making it impossible to observe labeling.

The bimolecular rate constant for labeling an exposed thiol at pH 8.5 is 7,000–8,000 M$^{-1}$·s$^{-1}$ and drops by about 10-fold for every unit decrease of the pH.\(^{10,13}\) Therefore, the intrinsic rate of labeling for 1 mM DTNB at pH 8 is about 1 s$^{-1}$.

Equilibrium Melts and Unfolding Rates. CD data were collected using an Aviv 410 spectropolarimeter with a Peltier temperature-controlled sample holder and an Aviv 62A DS spectropolarimeter with an 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 [Urea] at 27 °C. Samples were allowed to
equilibrate overnight, and the signal for each sample was averaged over 60 s. For unfolding rates, 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.

Fluorescence data were acquired on a Jvon Horiba Fluoromax-3 with excitation at 280 nm (1-nm slit width) and emission at 340 nm (2-nm slit width) with an integration time of 0.1 s at T = 27 °C.

Rates were measured on dilution to final urea concentrations from 4 to 6 M urea. The rate in the absence of urea was determined by fitting the log of the rates as a function of urea to a line and extrapolating back to 0 M urea. Error bars were obtained from bootstrapping.

References


**Supplemental Figures**

![Figure 2-S1](image1)

**Figure 2-S1. Residuals for the single exponential fit to Fig. 2C**
(the labeling reaction for L286C with 1 mM DTNB), showing that a single exponential fit is reasonable.

![Figure 2-S2](image2)

**Figure 2-S2. Thiol labeling of the second predicted hidden allosteric site**
(A) An example of the labeling of residue S203C at 1 mM DTNB. (B) The dependence of the labeling rate on the concentration of the labeling reagent (DTNB) with error bars representing the SD from three replicates.
<table>
<thead>
<tr>
<th>Mutant</th>
<th>Unfolding Rate (s$^{-1}$)</th>
<th>Labeling Rate (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>2.9E-6</td>
<td>0</td>
</tr>
<tr>
<td>L286</td>
<td>2.5E-5</td>
<td>5.9E-4 ± 5.7E-5</td>
</tr>
<tr>
<td>A232</td>
<td>1.6E-5</td>
<td>3.6E-3 ± 1.0E-3</td>
</tr>
<tr>
<td>A249</td>
<td>1.3E-5</td>
<td>6.8E-4 ± 1.5E-4</td>
</tr>
<tr>
<td>S203</td>
<td>2.7E-4</td>
<td>1.5E-2 ± 3.4E-3</td>
</tr>
</tbody>
</table>

Table 2-S1. Experimental unfolding rates monitored by CD and DTNB labeling rates  
Error bars for the labeling rates represent the SD across the three replicates for labeling with each of three DTNB concentrations (nine data points total).
Chapter 3:
Tracing a protein’s folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange
This article is under review at eLife, and I am a co-first author on this work.

3.1 Abstract

The conformations populated during protein folding have been studied for decades; yet, their evolutionary importance remains largely unexplored. Ancestral sequence reconstruction allows access to proteins across evolutionary time, and new methods such as pulsed-labeling hydrogen exchange coupled with mass spectrometry allow determination of folding intermediate structures at near amino-acid resolution. Here, we combine these techniques to monitor the folding of the ribonuclease H family along the evolutionary lineages of *T. thermophilus* and *E. coli* RNase H. All homologs and ancestral proteins studied populate a similar folding intermediate despite being separated by billions of years of evolution. Even though this conformation is conserved, the pathway leading to it has diverged over evolutionary time, and rational mutations can alter this trajectory. Our results demonstrate that evolutionary processes can affect the energy landscape to preserve or alter specific features of a protein’s folding pathway.

3.2 Introduction

Protein folding, the process by which an unfolded polypeptide chain navigates its energy landscape to achieve its native structure,\(^1,2\) can be defined by the partially folded conformations (intermediates) populated during this process. Such intermediates are key features of the landscape; they can facilitate folding, but they can also lead to misfolding and aggregation, resulting in a breakdown of proteostasis and disease.\(^3–5\) While identifying and characterizing these intermediates is critical to understanding and engineering a protein’s energy landscape, their transient nature and low populations present experimental challenges. Recent technological improvements in hydrogen exchange monitored by mass spectrometry (HX-MS) have provided access to the structural and temporal details of these folding intermediates at near-single amino-acid resolution.\(^6–11\) This pulsed-labeling HX-MS approach is particularly well suited to studies of multiple variants or families of proteins, as it does not require large amounts of purified protein or NMR assignments. Thus, pulsed-labeling HX-MS can be used to address long-standing questions in the field: How robust is a protein’s energy landscape to changes in the amino acid sequence, and how conserved is the folding trajectory over evolutionary time?

Ribonuclease HI (RNase H) is an ideal system to investigate protein folding over evolutionary time. RNase H from *E. coli*, ecRNH* (the asterisk denotes a cysteine-free variant of RNase H), is arguably one of the best-characterized proteins in terms of its folding pathway and energy landscape. Both stopped-flow ensemble studies and single-molecule optical trap experiments demonstrate that this protein populates a major obligate intermediate before the rate-limiting step in folding.\(^12–16\) A rare population of this intermediate can also be detected under native-state conditions.\(^17\) Several homologs of RNase H have also been studied, yielding insight into the folding trends of extant RNases H.\(^18–20\)
In addition to comparing the folding pathways of homologs, one can use a phylogenetic technique called ancestral sequence reconstruction (ASR) to access the evolutionary history of a protein family and study the properties of ancestral proteins.\textsuperscript{21,22} ASR has been applied to a variety of protein families and in addition to revealing the evolutionary history, these ancestral proteins can act as intermediates in sequence space to uncover mechanisms underlying protein properties.\textsuperscript{23–30} Recently, ancestral sequence reconstruction was applied to the RNase H family and the thermodynamic and kinetic properties of seven ancestral proteins connecting the lineages of \textit{E. coli} and \textit{T. thermophilus} RNase H (ecRNH* and ttRNH*) were characterized.\textsuperscript{31–33} Stopped-flow kinetics monitored by circular dichroism (CD) demonstrate that all seven ancestral proteins populate a folding intermediate before the rate-limiting step. Additionally, the folding and unfolding rates show notable trends along the phylogenetic lineages, and the presence of a folding intermediate plays an important role in modulating these evolutionary trends.\textsuperscript{32}

For ecRNH*, multiple methods have confirmed the structural details of the folding intermediates. This major folding intermediate, termed I\textsubscript{core}, which forms before the rate limiting step, involves secondary structure from the core region of the protein, including Helices A-D and Strands 4 and 5, while the rest of the protein (Helix E and Strands 1, 2, 3), remains unfolded (Figure 3-1A).\textsuperscript{12,34,35} Pulsed-labeling HX-MS with near amino acid resolution was developed using ecRNH* as the model protein.\textsuperscript{6} This approach confirmed the structure of I\textsubscript{core} and revealed the stepwise protection of individual helices leading up to the intermediate. Specifically, the amide hydrogens in Helix A and Strand 4 are the first elements to gain protection, followed by those in Helix D and Strand 5,

![Figure 3-1. RNase H structure and Pulsed-labeling HX-MS](image)

\textbf{A} Crystal structure of \textit{E. coli} RNase H* (ecRNH*) (PDB: 2RN2).\textsuperscript{55} Secondary structural elements: Red: Strand 1, Strand 2, Strand 3 (S123); Blue: Helix A, Strand 4 (HAS4); Yellow: Helix B, Helix C (HBHC); Green: Helix D, Strand 5 (HDSS); Purple: Helix E (HE). The core region of the protein (I\textsubscript{core}) involving Helix A, Strand 4, Helix B, Helix C, Helix D, Strand 5 and the periphery region of the protein involving Strand 1, Strand 2, Strand 3, Helix E are denoted. \textbf{B} Pulsed-labeling setup and workflow. Unfolded, fully deuterated protein in high [urea] is rapidly mixed with low [urea] refolding buffer to initiate refolding. After some refolding time, hydrogen exchange of unprotected amides is initiated by mixing with high-pH pulse buffer. The hydrogen exchange reaction is quenched by mixing with a low-pH quench buffer. The sample is injected onto an LC-MS for in-line proteolysis, desalting, and peptide separation by reverse-phase chromatography followed by MS analysis.
and then Helices B and C to form the canonical I\textsubscript{core} intermediate. The periphery, comprising of Strands 1-3 and Helix E, gains protection in the rate-limiting step to the native state. Would this I\textsubscript{core} folding intermediate and the stepwise folding pathway be conserved across evolution?

Here, we use pulsed-labeling HX-MS on the resurrected family of RNases H to investigate the evolutionary and sequence determinants governing the folding trajectory. Specifically, we find that the structure of the major folding intermediate (I\textsubscript{core}) has been conserved over three billion years of evolution, suggesting that this partially folded state plays a crucial role in the folding or function of the protein. The detailed steps leading to this folding intermediate, however, vary. The very first step in folding differs between the two extant homologs: for ecRNH\textsuperscript{*}, Helix A gains protection before Helix D, while for ttRNH\textsuperscript{*}, Helix D acquires protection before Helix A. This pattern can be followed along the evolutionary lineages: most of the ancestors fold like ttRNH\textsuperscript{*} (Helix D before Helix A) and a switch to fold like ecRNH\textsuperscript{*} (Helix A before Helix D) occurs late along the mesophilic lineage. These phylogenetic trends allow us to investigate how these early folding events are encoded in the amino acid sequence. By selectively modulating biophysical properties, notably intrinsic helicity, of specific secondary structure elements, we are able to favor or disfavor the formation of specific conformations during folding and have engineering control over the folding pathway of RNase H.

3.3 Results

Monitoring a protein’s folding trajectory by pulsed-labeling HX-MS

We used pulsed-labeling hydrogen exchange monitored by mass spectrometry (HX-MS) on extant, ancestral, and site-directed variants of RNase H to examine the robustness of a protein’s folding pathway to sequence changes. These experiments allow us to characterize the partially folded intermediates and the order of structure formation during folding to ask whether these intermediates have changed over evolutionary time, and what role sequence might play in determining these intermediates.

Figure 1B outlines the scheme for the pulsed-labeling experiment (for details, see Methods). Briefly, folding is initiated by rapidly diluting an unfolded (high [urea]), fully deuterated protein into folding conditions (low [urea]) at 10°C. After various folding times (t\textsubscript{f}), a pulse of hydrogen exchange is applied to label amides in regions that have not yet folded. The amount of exchange at each folding timepoint is then detected by in-line proteolysis and LC/MS. Data are analyzed first at the peptide level by monitoring the protection of deuterons on peptides as a function of refolding time, and then at the residue level, using overlapping peptides de-convoluted by the program HDsite.\textsuperscript{36,37}

Since the original folding studies on RNase H were carried out at 25°C, we re-characterized the folding of each RNase H variant at 10°C using stopped-flow circular dichroism spectroscopy (Figure 3-S1). The refolding profiles were consistent with those at 25°C.\textsuperscript{12,19,32} At low [urea], all ancestors show a large signal change (burst phase) within the dead time of the stopped-flow instrument (~15 ms), followed by a slower observable phase which fit well to a single exponential. The resulting chevron plots
(ln(\(k_{\text{obs}}\)) vs [urea]) show the classic rollover at low [urea] due to the presence of a stable folding intermediate. As expected, the observed rates at 10°C are slower than 25°C, but the chevron profiles are similar for all RNase H variants. Thus, the overall folding trajectory, notably the population of a folding intermediate, has not changed between the two temperatures.

**Monitoring the folding pathway of ttRNH* using pulsed-labeling HX-MS**

First, we characterized the conformations populated during folding of extant RNase H from *T. thermophilus* and compared its folding trajectory to the previously characterized folding trajectory of *E. coli* RNase H. 374 unique peptides were identified by MS. Of these, 49 unique peptides were observed at all refolding time points and were used for further analysis (Figure 3-2A). Similar to ecRNH*, peptides associated with I_{core} (Helix A-D, Strands 4-5) gain protection early (within milliseconds), corresponding to the timescale for the formation of the folding intermediate. Peptides associated with the periphery of the protein (Strands 2-3, Helix E) gain protection on the order of seconds, corresponding to the rate-limiting step (Figure 3-2B). Thus, the major folding intermediate in ttRNH*, I_{core}, is strikingly similar to that of ecRNH*.

Looking at the very early refolding times allows one to determine the individual folding steps preceding I_{core}. At the earliest time point (~1 ms), almost all peptides are unfolded (fully exchange with solvent) with the exception of those in Helix D and Strand 5, which are ~40% deuterated (Figure 3-2C). Peptides spanning Helix A and Strand 4 are less protected (~15% deuterated) at this same time point. This order of protection (Helix D before Helix A) is notably different than that for *E. coli* RNase H*, where Helix A is protected before Helix D. Peptides spanning Helix B and Helix C gain protection in the I_{core} intermediate. Peptides from Strands 1-3 and Helix E do not gain full protection until significantly later (on the order of seconds), corresponding to the rate-limiting step to the native state. Thus, while the I_{core} intermediate is largely conserved between ttRNH* and ecRNH*, the initial steps of folding differ between the two homologs.

The peptide data from each time point were also analyzed using HDSite to determine residue-level protection in a near site-resolved manner (Figure 3-2D). These site-resolved data also show protection appearing first in Helix D and Strand 5, followed by Helix A/Strand 4, Helix B/C, and finally, the periphery Helix E and Strands 1-3. The differences in the order of protection leading up to I_{core} of ecRNH* and ttRNH* are also evident in this site-resolved analysis.

**Pulsed-labeling HX-MS on the ancestral RNases H**

To look for evolutionary trends in the folding trajectory, we probed the folding pathway of ancestral RNases H along the lineages of *E. coli* and *T. thermophilus* RNase H (Figure 3-3A). Anc1* is the last common ancestor of ecRNH* and ttRNH*. Anc2* and Anc3* are ancestors along the thermophilic lineage leading to ttRNH*, and AncA*, AncB*, AncC*, and AncD* are ancestors along the mesophilic lineage leading to ecRNH*. Previous kinetic studies demonstrated that all of the ancestral proteins fold via a three-state pathway, populating an intermediate before the rate-limiting step. We
Figure 3-2. Determination of the folding pathway of *T. thermophilus* RNase H* by HX-MS

(A) Protection of representative peptides from ttRNH* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel B. The dotted arrow indicates the refolding time point analyzed in panel C. (B) Protection of peptides mapping to the core region (I\textsubscript{core}) or the periphery region of ttRNH* at 21 msec after refolding. Bars represent the mean and standard deviation of each data set. *p < 0.0001 (Welch’s unpaired T-test).

(C) Protection of peptides of ttRNH* mapping to distinct secondary structural elements at 1 msec after refolding. Bars represent the mean and standard deviation of each data set. *p = 0.0027 (Welch’s unpaired T-test).

(D) Residue-resolved folding pathway of ttRNH* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with a “x”.

**ttRNH* Peptides**
- 1-11
- 1-29
- 12-26
- 113-116
- 54-71
- 142-166
- 56-71
- 146-161
- 62-71
- 146-157
- 72-80
- 72-86
- 72-89
now use pulsed-labeling HX-MS to obtain a near-site resolved trajectory of the folding pathway for each ancestor and determine whether the $I_{\text{core}}$ structure is conserved over evolution.

We obtained good peptide coverage for all of the ancestors with a minimum of 81 peptides seen in all time points for each variant (Figure 3-3, Figures 3-S2-S7). As observed in both ttRNH* (above) and ecRNH* all of the ancestral RNases H populate the canonical $I_{\text{core}}$ folding intermediate prior to the rate-limiting step. Peptides corresponding to the $I_{\text{core}}$ region of the RNase H structure become protected on the timescale of milliseconds, while the rest of the protein gains protection on the timescale of seconds (Figure 3-3C, Figures 3-S2-S7). Thus, the structure of this major folding intermediate is not only present in both extant RNases H but is conserved over nearly three billion years of evolutionary history.

Similarly, to the extant proteins, the periphery of the ancestral proteins gains protection on a much slower timescale (Figure 3-3C, Figure 3-S2-S7). The details of protection in this region, however, vary somewhat across the ancestors. The periphery becomes fully protected by the last time point in all ancestral proteins except for AncB* (Figure 3-S5). AncB* was previously characterized to be non-two-state with a notable population of the folding intermediate under equilibrium conditions, and the lack of protection in the periphery in the folded state of AncB* is consistent with this observation. For Anc1* and Anc2*, there are also notable differences in the time course of protection for the terminal helix, Helix E. For these two proteins, the peptides spanning Helix E are decoupled from Strands 1-3 (which show protection on the same timescale as global folding) and do not gain protection even in the folded state of the protein (Figure 3-3B, Figure 3-3D, Figure 3-S2), suggesting that Helix E is improperly docked or poorly structured in Anc1* and Anc2*. Indeed, Helix E is known to be labile in ecRNH*: a deletion variant of ecRNH* without this final helix forms a cooperatively folded protein, and recent single-molecule force spectroscopy of ecRNH* showed that Helix E can be pulled off the folded protein under low force while the remainder of the protein remains structured (manuscript in preparation). It appears that Helix E may be further destabilized in Anc1* and Anc2* such that it does not show protection in the native state.

The early folding steps of RNase H change across evolutionary time

Since the order of events leading to $I_{\text{core}}$ differs between the extant homologs, we examined whether the ancestral RNases H spanning the lineages of these two homologs show any trends in their early folding steps. For each ancestor, we analyzed the fraction of deuterium protected in peptides that are uniquely associated with specific helices of the protein (Figure 3-3D and Figure 3-S2-S7) to determine which regions fold first.

These data show that the last common ancestor of ecRNH* and ttRNH*, Anc1*, as well as all proteins along the thermophilic lineage (Anc2* and Anc3*) show similar behavior to ttRNH* and gain protection first in Helix D/Strand 5 (Figure 3-S2, 3-S3). For the first two ancestors along the mesophilic lineage (AncA* and AncB*), the order of
Figure 3-3. Determination of the folding pathway of ancestral RNases H by HX-MS

(A) Representation of the phylogenetic tree of the RNase H family illustrating the ancestral proteins along the two lineages leading to *E. coli* RNase H and *T. thermophilus* RNase H. Adapted from Figure 2A of Hart KM et al. 2014, *PLoS Biology* 12(11) doi:10.1371/journal.pbio.1001994, published under the Creative Commons Attribution 4.0 International Public License (CC BY 4.0; https://creativecommons.org/licenses/by/4.0/). Anc1* is the last common ancestor of ecRNH* and ttRNH*. Anc2* and Anc3* are ancestors along the thermophilic lineage to ttRNH*. AncA*, AncB*, AncC*, and AncD* are ancestors along the mesophilic lineage to ecRNH*. (B) Protection of representative peptides from Anc1* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel C. The dotted arrow indicates the refolding time point analyzed in panel D. (C) Protection of peptides mapping to the core region (I<sub>core</sub>) or the periphery region of Anc1* at 13 msec after refolding. Bars represent the mean and standard deviation of each data set. *p = 0.0011 (Welch’s unpaired T-test) (D) Protection of peptides mapping to distinct secondary structural elements of Anc1* at 1 milliseconds after refolding. Bars represent the mean and standard deviation of each data set. *p < 0.0001 (Welch’s unpaired T-test). (E) Residue-resolved folding pathway of Anc1* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with a "x".
protection is difficult to determine. For AncA*, there is no significant difference in the degree of protection among the peptides within I\textsubscript{core} (this analysis is limited by the availability of peptides associated exclusively within a region) (Figure 3-S4). However, when all overlapping peptides are analyzed using HDSite to obtain site resolution, we observe notable protection in Helix D at the earliest refolding times. Therefore, we conclude that although Helix D folding before Helix A is likely, the early folding events of AncA* cannot be unambiguously determined. For AncB*, all of I\textsubscript{core} gains protection at

**Figure 3-4. Intrinsic helicity as a predictor for the early folding mechanism of RNases H**

Log-ratio of intrinsic helicity of Helix A and Helix D for each RNase H variant studied. Intrinsic helix predictions were calculated using AGADIR.\textsuperscript{39} The order of helix protection for each variant of RNase H is depicted in color. Green bars represent proteins where Helix D is the first structural element to gain protection during refolding. Blue bars represent proteins where Helix A is the first structural element to gain protection during refolding. Grey bars represent proteins where the helix protection order could not be unambiguously determined. The order of helix protection for each ancestor and homolog is also colored on the phylogenetic tree, revealing a trend in the RNase H folding trajectory along the evolutionary lineages. The phylogenetic tree shown in this figure is adapted from Figure 2A of Hart KM et al. 2014, *PLoS Biology*, 12(11) doi:10.1371/journal.pbio.1001994, published under the CreativeCommons Attribution 4.0 International Public License (CC BY 4.0; https://creativecommons.org/licenses/by/4.0/).\textsuperscript{31}
the same time point, both at the peptide and residue-level, so the order of assembly cannot be determined with our time resolution (Figure 3-S5).

The next ancestor along the mesophilic lineage, AncC*, shows protection first in Helix D, indicating that this pattern of protection is maintained through the mesophilic lineage to this ancestor (Figure 3-S6). AncD*, the most recent ancestor along the mesophilic lineage, however, is similar to ecRNH* and gains protection first in Helix A (Figure 3-S7). As detailed for the other ancestors, the data were also analyzed using HDSite to determine residue-level protection for each ancestral RNase H (Figure 3-3E, Figure 3-S2-S7). These data indicate a pattern in the order of protection in the early steps of the folding pathway across the RNase H ancestors. Early protection in Helix D is an ancestral feature of RNase H that is maintained in the thermophilic lineage, with a transition occurring late during the mesophilic lineage to a different pathway where Helix A is protected before Helix D, resulting in a distinct folding pathway for the two extant RNase H homologs (Figure 3-4).

**Early helix protection is determined by the local sequence of the core**

Relative to the vast sequence space available, these RNase H ancestors represent a set of closely related sequences with distinct folding properties and provide an excellent system to help us elucidate the physiochemical mechanism and the sequence determinants dictating the RNase H folding trajectory. An analysis of the intrinsic helical propensity of each region using the algorithm AGADIR shows a notable trend in helicity that correlates with the early folding events (Figure 3-4). For proteins that gain protection in Helix A first, the intrinsic helicity of Helix A is four-fold higher than that of Helix D. For the variants where Helix D is protected first, the intrinsic helicity of Helix D is similar to or greater than Helix A. This suggests that intrinsic helix propensity may play an important role in determining which region is the first to gain protection during the folding pathway of RNase H. To investigate this hypothesis, we turned to rationally designed variants.

**Intrinsic helicity plays a role in determining the structure of the early intermediates**

If the order of protection in the early folding events of RNase H is determined by intrinsic helix propensity, then we should be able to alter the protein sequence rationally and manipulate the folding trajectory. Thus, we asked whether single-site mutations that change the relative helix propensity of Helix A and Helix D could alter the folding trajectory of ecRNH* and make it fold in a similar fashion to trRNH*. Two different point mutations were made in ecRNH*: A55G decreases helix propensity in Helix A, and D108L increases helicity in Helix D (Figure 3-4, Figure 3-5A, Table 3-S1). Pulsed-labeling HX-MS indicates that both of these variants alter the early folding events of ecRNH*. The peptide-level protection of ecRNH* A55G indicates that at 13 ms, both Helix A and Helix D show similar levels of protection. In contrast, for wild-type ecRNH*, Helix A shows protection by 1 ms and Helix D does not show comparable protection until 10-20 msec. Thus, the mutation A55G slows the gain of protection in Helix A such
that it no longer protected before Helix D (Figure 3-5B). The peptide-level protection of ecRNH* D108L indicates a change in the order of protection. Due to the limited number of peptides available, we could only confidently determine this using peptides spanning the N-terminus of Helix D. At 13 ms, the N-terminus of Helix D (residues 106-108) near the D108L mutation is protected significantly faster than any other region of the protein. Thus increasing helix propensity correlated with a change in the folding trajectory. (Figure 3-5C). Together, these two mutations suggest that intrinsic helicity plays a role in the early folding events of RNase H and can be used to alter the stepwise order of conformations populated during folding.

Figure 3-5. Engineered mutations to alter the folding pathway of ecRNH*
(A) Crystal structure of E. coli RNase H (PDB: 2RN2) with mutations designed to alter intrinsic helicity. A55G, located in Helix A (blue), is colored in cyan. D108L, located in Helix D (green), is colored in light green. (B) Protection of peptides mapping to distinct secondary structural elements of ecRNH* A55G at 13 msec after refolding. Bars represent the mean and standard deviation of each data set. p = 0.0917 (n.s. = not significant, Welch’s unpaired T-test). (C) Protection of peptides mapping to distinct secondary structural elements of ecRNH* D108L at 13 msec after refolding. Bars represent the mean and standard deviation of each data set. *p = 0.0016, **p = 0.0044 (Welch’s unpaired T-test)

3.4 Discussion

Determining the folding pathway of multiple protein variants

Pulsed-labeling hydrogen exchange is currently the most detailed method to identify the conformations populated during protein folding. This approach was initially developed for use with NMR detection where it benefited from NMR’s site-specific resolution of individual amides. However, using NMR with pulsed-labeling HX requires tens of milligrams of sample and NMR peak assignments for the amides in each protein studied. In addition, probes are limited to amide sites stable to exchange in the final folded state (protection factors of >~80,000) resulting in loss of information at individual sites, which can sometimes represent large regions of the protein. In contrast, detection by mass spectrometry as applied in this study requires much less protein sample, has much faster data collection, and can theoretically cover 100% of the protein sequence. Importantly, this approach does not demand any structural information of the folded state, such as NMR assignments, for the specific protein or variant studied. These advantages enabled us to obtain the stepwise folding pathway of nine variants of RNase H and study the evolutionary history and sequence determinants of the RNase H folding pathway in detail. While pulsed-labeling HX-MS has been used to characterize
the folding pathways of several model systems, this study is the first to utilize the higher throughput nature of HX-MS to study an ensemble of protein variants. The advantages of this technique to study many different sequences of the same fold shows great promise for probing the relationship between amino acid sequence and a protein’s energy landscape and will likely be particularly valuable for protein engineering and design applications.

**Icore is a structurally conserved folding intermediate over 3 billion years of evolution**

The native fold of a protein is robust to changes in sequence, proteins with >~30% sequence identity share the same fold. Thus small variations in sequence, such as those found among homologs or site-specific mutations, do not affect the overall three-dimensional structure of a protein. These mutations can, however, affect the overall energy landscape, which in turn can have profound effects of function. Here, we find conservation of a high-energy structure populated during the folding of the RNase H family over incredibly long evolutionary timescales. Using pulsed-labeling HX-MS we identified and characterized the structure of the major folding intermediate in seven ancestral and several mutant RNases H, which together with previous studies on extant homologs, suggest that the conservation of this intermediate is a key feature of the RNase H energy landscape across ~3 billion years of evolutionary time.

Why does Icore persist on the energy landscape of RNase H? One explanation is a simple topological constraint; all RNases H may need to fold via a populated Icore intermediate to successfully reach the native state. This explanation, however, is countered by a previous study where a single mutation (I53D) in ecRNH* destabilizes Icore such that it is no longer significantly populated during folding—yet this variant still folds to the native state. Adding osmolytes, such as sodium sulfate, stabilizes this folding intermediate and switches ecRNH* I53D back to a three-state folding pathway, showing that the presence of the folding intermediate can be modulated. Additionally, a fragment of RNase H containing only the Icore sequence (and variants thereof) can autonomously fold and be studied at equilibrium, indicating that this structure is stable and robust to mutations. The nature of the rate-limiting step, or folding barrier, which allows for the buildup of this intermediate is unclear. One possibility is that the Icore intermediate is populated simply because the information for folding this region is completely encoded locally and Icore can fold relatively fast, before this rate limiting step to the fully folded state.

Alternatively, Icore could be conserved because it contributes to the biological function or fitness of the protein. Partially folded states and high-energy non-native conformations are known to be important for a variety of protein functions and proteostasis. All of the ancestral RNases H we studied here are active, in that they cleave RNA-DNA hybrids in vitro; and although the residues thought to contribute to substrate-binding affinity are contained in the core region of the protein, the active site residues (D10, E48, D70) span both the core and the periphery. It is therefore possible
that a stable folding core with an energetically independent periphery is important for the efficiency or dynamics associated with catalysis in RNase H.

While the presence of the I\textsubscript{core} intermediate has been observed in all proteins studied here, recent studies have suggested that some of the RNase H variants, notably for proteins along the thermophilic lineage, the I\textsubscript{core} folding intermediate may also involve structure in the first β-strand.\textsuperscript{33,43,47} While we see slight protection in this region for ttRNH\textsuperscript{*}, hydrogen exchange may not be the best probe of this—docking of Strand 1 without its hydrogen-bonding partners in the rest of the β-sheet may not be reflected by backbone amide protection. Therefore, amide protection may not be observed even if Strand 1 docks early to the core. The involvement of Strand 1 in ancestral other RNase H variants studied remains unclear from this study.\textsuperscript{33,43}

**Aspects of the folding pathway are malleable across evolutionary time**

Our pulsed-labeling HX-MS results also illustrate how other features of a protein’s energy landscape can be altered over evolutionary timescales. Although the I\textsubscript{core} intermediate is conserved across all RNases H studied, the individual folding steps leading up to I\textsubscript{core} differ. Anc1\textsuperscript{*}, the last common ancestor, folds through a pathway where the Helix D/Strand 5 region is the first structural element to gain protection. This ancestral feature is maintained along the thermophilic lineage to the extant ttRNH\textsuperscript{*}. Along the mesophilic branch, we observe a switch from this ancient folding pathway to one that first forms protection in Helix A/Strand 4 that occurs evolutionarily between AncC\textsuperscript{*} and AncD\textsuperscript{*}. This suggests that while the structure of I\textsubscript{core} has been conserved across 3 billion years of evolution, the steps to form this intermediate are malleable over time. Since an isolated helix is unlikely show protection by HX, we expect additional hydrophobic collapse of the polypeptide to contribute to the observed protection. Nonetheless, the switch in protection between Helix A and Helix D indicates that formation of native structure nucleates in a different region of the protein across the RNase H variants studied, with a clear evolutionary trend.

Despite these trends, it remains difficult to rationalize these observations in terms of a selective evolutionary pressure or fitness implication. These very early events occur on the order of one millisecond, significantly faster than the overall folding of the protein. Furthermore, all of these RNase H proteins fold to their native state efficiently with no evidence for aggregation or misfolding. So, although partially folded states have been implicated as gateways for aggregation for some proteins,\textsuperscript{4} this does not appear to be the case for RNase H. It is possible that the change in the early folding step is a result of mutations that are coupled to another feature under selection or drift. Although the actual evolutionary implication for the RNase H folding pathway may be lost in history, the trend in folding pathway across evolutionary time demonstrates that folding pathways and conformations on the energy landscape of proteins can be affected over time, and this system provides an excellent tool to interrogate the role sequence plays in guiding the process of protein folding.
The folding pathway of RNase H can be altered using simple sequence changes

Our study also shows how insights from evolutionary history can contribute to our understanding of the physiochemical mechanisms dictating the protein energy landscape and how we might use that knowledge to engineer the landscape. The regions that gain protection first involve helical secondary structure elements, and their folding order correlates with isolated helical propensity of these regions predicted by AGADIR. Proteins where protection is first observed in Helix A have higher intrinsic helicity in Helix A than in Helix D. Proteins where Helix D gains protection first have higher helicity in Helix D or roughly equal helicity in both regions. This property was used to guide our site-directed mutagenesis to select variants to alter the folding trajectory of ecRNH* in a predictive manner using intrinsic helicity as a guide.

While these results are consistent with local helicity as a determinant of the earliest folding steps, there may be other parameters that dictate the formation of these conformations. The parameter average area buried upon folding (AABUF) which measures the average change in surface area of a residue from an unfolded state to a folded state, has been shown to correlate to the structure of the folding intermediate in apomyoglobin. Both helicity and AABUF are altered in the mutants considered in our study (Table 3-S1). Indeed, AABUF and helicity are often correlated and contributions of either parameter are difficult to disentangle. Nevertheless, our data suggest that parameters that are locally encoded in regions of a protein can be used engineer the energy landscape of a protein including its folding pathway.

We have used a combination of ASR and pulsed-labeling HX-MS to explore the conformations populated during the folding of multiple RNase H proteins, including homologs, ancestors, and single-site variants. All RNase H proteins studied populate the same major folding intermediate, I\text{core}, indicating that this conformation has been maintained on the energy landscape of RNase H over long evolutionary timescales (>3 billion years). This remarkable conservation of a partially folded structure on the energy landscape of RNase H is contrasted with changes in the folding pathway leading up to this structure. The early folding events preceding this intermediate (Helix A protected before Helix D or vice versa) differ between the two homologs and also shows a notable trend along the evolutionary lineages. This pattern of protection correlates with the relative helix propensity of the sequences comprising these two helices, and we use this knowledge to alter the folding pathway of ecRNH* through rationally designed mutations. Our study illustrates how the energy landscape of a protein can be altered in complex ways over evolutionary time scales, and how insights from evolutionary history can contribute to our understanding of the physiochemical mechanisms dictating the protein energy landscape.

3.5 Materials and Methods

Protein Purification
Cysteine-free \textit{T. thermophilus} RNase H, and ancestral RNases H were expressed and purified as previously described. Point mutants were generated using site-directed mutagenesis, confirmed by Sanger sequencing, and the proteins
were purified as previously described.\textsuperscript{53} Purity was confirmed by SDS-PAGE and mass spectrometry.

**HX-MS System**

Hydrogen exchange mass spectrometry (HX-MS) experiments were carried out using a system similar to that described by Mayne \textit{et al.}\textsuperscript{7,8} Briefly, a Bio-Logic SFM-4/Q quench flow mixer with a modified head piece with reduced swept volume was used to initiate protein refolding, followed by pulse-labeling unprotected amide hydrogen atoms, and quenching of the labeling reaction. The minimum dead time for mixing is 13 msec. Quenched samples were injected into an HPLC system constructed using two Agilent 1100 HPLC instruments. The quenched sample was flowed over columns (Upchurch C130B) packed with beads of immobilized pepsin and fungal protease at 400 μL/min in 0.05% TFA. The digested protein was run onto a C-4 trap column (Upchurch C-128 with POROS R2 beads) for desalting. An acetonitrile gradient (15-100% acetonitrile, 0.05% TFA at 17 μL/min) eluted peptides from this C-4 trap column and onto an analytical C-8 column (Thermo 72205-050565) for separation before injection into an ESI source for mass spectrometry analysis on a Thermo Scientific LTQ Orbitrap Discovery. The entire HPLC system is kept submerged in an ice bath at 0°C to reduce back exchange of deuterium atoms during the chromatography steps. The workflow takes ~10-18 minutes from injection to peptide detection.

**Refolding Experiment**

Similar to previous reports,\textsuperscript{6,8} unfolded protein samples in high denaturant (80 μM [protein], 20 mM NaOAc pH=4.1, 7-9 M [urea]) were deuterated by a repeated cycle of lyophilization and resuspension in D$_2$O. For the pulsed labeling experiment, 1 volume of deuterated protein was mixed in the SFM-4/Q with 10 volumes of refolding buffer (10 mM Sodium Acetate pH=5.29, H$_2$O) to initiate refolding. The pulse for hydrogen exchange was initiated by mixing with 5 volumes of high pH buffer (100 mM Glycine pH=10.11) and then quenched after 10 ms with 5 volumes low pH buffer (200 mM Glycine pH=1.95). The length of the delay line between the first and second mixer was changed to achieve a range of refolding times. An interrupted mixing protocol was used to measure the longest refolding time points (>373 ms). Undeuterated protein was used to perform tandem mass spectrometry (MS/MS) analysis to compile a list of peptides and their retention times in the HPLC system. Competition experiments where refolding and exchange were initiated at the same time were performed by diluting deuterated protein in high urea into high-pH refolding buffer (100 mM Glycine pH=10.11). In this experiment each site will exchange with the solvent around it unless it can gain protection before exchange occurs (<1 ms on average). For each time point, an identical sample was collected in which the high pH pulse was replaced by unbuffered water to measure back exchange for each sample. All data were obtained in triplicate and were normalized for back exchange. Data for ttRNH$^*$ were normalized to the theoretical maximum number of deuterons as back exchange controls for this protein did not produce enough peptides. Fully folded controls were created by diluting unfolded
protein samples 1:10 in fully deuterated refolding buffer and incubating at room temperature for 4 hours before applying the same 10 ms high-pH pulse using the SFM-4/Q.

**MS detection and data analysis**

Proteome Discoverer 2.0 (Thermo Scientific) was used to identify peptides from the tandem MS data. Peptides identified in the pulse-labeled refolding experiments with deuterated protein were used to determine the presence and deuteration level of each peptide at each refolding time point. The spectral envelope of each peptide was fit using two separate algorithms developed by the Englander Lab to determine their deuteration state — ExMS for identification and fitting of peptides and HDSite for deconvolution of overlapping peptides to achieve near-amino acid level deuteration levels.\(^{36,37}\) In addition, HD Examiner (Sierra Analytics) was used to identify and fit each peptide and determine deuteration levels. Different charge states of the same peptide were averaged where noted and used for further analysis. Centroids of each peptide at each time point taken from HD Examiner were used for further analysis. The residue cutoffs for specific structural regions of each protein were determined from a multiple sequence alignment using the structure of *E. coli* RNase H as a guide (PDB: 2RN2).\(^{31}\) Peptides were assigned to different structural regions based on these residue cutoffs. Peptides that spanned multiple secondary structural regions of a protein were excluded from further analysis, as were peptides not present in all time points. Peptides mapping to Strands 1-3 and Helix E were assigned to the periphery region of the protein. Peptides mapping to Helix A-D and Strands 4-5 were assigned to the core region of the protein.

**3.6 References**


50. Nishimura, C., Dyson, H. J. & Wright, P. E. Consequences of Stabilizing the


3.7 Supplemental Materials

Figure 3-S1. Chevron plot of RNase H variants studied at 10°C and 25°C
Chevron plots \((\ln(k_{obs})\text{ vs } [\text{urea}])\), determined from refolding and unfolding experiments in various [urea] at 10°C and 25°C for (A) ttRNH*, (B) Anc1*, (C) Anc2*, (D) Anc3*, (E) AncA*, (F) AncB*, (G) AncC*, (H) AncD*, (I) ecRNH* A55G, (J) ecRNH* D108L. For (A) Both the fast (black dots) and slow (grey dots) rates of folding for ttRNH* are shown at 10°C, and chevron fits for the two rates at 25°C are shown as lines and adapted from previous work.\(^{19}\) Data at 25°C for (B) – (H) were adapted from a previously published study.\(^{32}\)
Figure 3-S2.
Determination of the folding pathway of Anc2* by HX-MS

(A) Protection of representative peptides from Anc2* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel B. The dotted arrow indicates the refolding time point analyzed in panel C. (B) Protection of peptides mapping to the core region (I_{core}) or the periphery region of Anc2* at 21 msec after refolding. Bars represent the mean and standard deviation of each data set. *p = 0.0011 (Welch’s unpaired T-test) (C) Protection of peptides of Anc2* mapping to distinct secondary structural elements at 1 msec after refolding. Bars represent the mean and standard deviation of each data set. *p = 0.0064 (Welch’s unpaired T-test). (D) Residue-resolved folding pathway of Anc2* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with a “x”. 
Figure 3-S3. Determination of the folding pathway of Anc3* by HX-MS
(A) Protection of representative peptides from Anc3* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel B. The dotted arrow indicates the refolding time point analyzed in panel C. (B) Protection of peptides mapping to the core region (I_{core}) or the periphery region of Anc3* at 13 msec after refolding. Bars represent the mean and standard deviation of each data set. *p < 0.0001 (Welch’s unpaired T-test). (C) Protection of peptides of Anc3* mapping to distinct secondary structural elements at 1 msec after refolding. Bars represent the mean and standard deviation of each data set. *p = 0.0419 (Welch’s unpaired T-test). (D) Residue-resolved folding pathway of Anc3* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with an “x”.

Anc3* Peptides
- 3-11: 97-112
- 11-29: 99-112
- 11-37: 100-112
- 51-61: 138-162
- 55-71: 141-162
- 62-74: 145-162
- 72-80
- 81-93
- 82-92
Figure 3-S4. Determination of the folding pathway of AncA* by HX-MS

(A) Protection of representative peptides from AncA* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel B. The dotted arrow indicates the refolding time point analyzed in panel C. (B) Protection of peptides mapping to the core region (I_{core}) or the periphery region of AncA* at 373 msec after refolding. Bars represent the mean and standard deviation of each data set. *p < 0.0001 (Welch’s unpaired T-test) (C) Protection of peptides of AncA* mapping to distinct secondary structural elements at 40 msec after refolding. Bars represent the mean and standard deviation of each data set. p = 0.275 (n.s. = not significant, Welch’s unpaired T-test). (D) Residue-resolved folding pathway of AncA* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with a “x”.

AncA* Peptides
- 8-23
- 8-25
- 24-32
- 112-123
- 114-123
- 51-68
- 68-131
- 68-134
- 56-68
- 59-68
- 141-155
- 72-82
- 74-82
- 74-88

<table>
<thead>
<tr>
<th>Refolding Time (sec)</th>
<th>Normalized Fraction Deuterated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>0.2</td>
</tr>
<tr>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>10.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Panel B:
- Periphery
- Core

Panel C:
- Strands 1-3
- Helix A
- Helix B
- Helix C
- Helix D
- Helix E

Panel D:
- Competition <1 msec
- 40 msec
- 377 msec
- 1 sec
- 110 sec
- Folded
Figure 3-S5. Determination of the folding pathway of AncB* by HX-MS

(A) Protection of representative peptides from AncB* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel B. The dotted arrow indicates the refolding time point analyzed in panel C. (B) Protection of peptides mapping to the core region (I_{core}) or the periphery region of AncB* at 21 msec after refolding. Bars represent the mean and standard deviation of each data set. *p < 0.0001 (Welch’s unpaired T-test) (C) Protection of peptides of AncB* mapping to distinct secondary structural elements at 1 msec after refolding. Bars represent the mean and standard deviation of each data set. p = 0.353 (n.s. = not significant, Welch’s unpaired T-test). (D) Residue-resolved folding pathway of AncB* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with a “x”.

AncB* Peptides
- 5-26
- 7-25
- 26-33
- 51-64
- 53-67
- 57-67
- 71-87
- 73-86
- 77-85

AncB* - 21 msec

AncB* - 1 msec
**Figure 3-S6.** Determination of the folding pathway of AncC* by HX-MS

(A) Protection of representative peptides from AncC* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel B. The dotted arrow indicates the refolding time point analyzed in panel C. (B) Protection of peptides mapping to the core region (I\text{core}) or the periphery region of AncC* at 77 msec after refolding. Bars represent the mean and standard deviation of each data set. *p < 0.0001 (Welch's unpaired T-test) (C) Protection of peptides of AncC* mapping to distinct secondary structural elements at 13 msec after refolding. Bars represent the mean and standard deviation of each data set. *p<0.0001 (Welch's unpaired T-test). (D) Residue-resolved folding pathway of AncC* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with "x".
Determination of the folding pathway of AncD* by HX-MS

(A) Protection of representative peptides from AncD* at various refolding times. Peptides are colored according to their corresponding structural element. The solid arrow indicates the refolding time point analyzed in panel B. The dotted arrow indicates the refolding time point analyzed in panel C. (B) Protection of peptides mapping to the core region (I_{core}) or the periphery region of AncD* at 40 msec after refolding. Bars represent the mean and standard deviation of each data set. *p < 0.0001 (Welch's unpaired T-test). (C) Protection of peptides of AncD* mapping to distinct secondary structural elements at 13 msec after refolding. Bars represent the mean and standard deviation of each data set. *p = 0.021 (Welch's unpaired T-test). (D) Residue-resolved folding pathway of AncD* at representative refolding time points. Data points in black indicate residues that are site-resolved. Data points in grey indicate residues in regions with less peptide coverage and are thus not site-resolved with the neighboring residues. Residues where site-resolved protection could not be determined due to insufficient peptide coverage is denoted with a “x”.

Figure 3-S7.
Table S1. Comparison of intrinsic helicity and AABUF across RNase H variants

<table>
<thead>
<tr>
<th></th>
<th>tlRNH*</th>
<th>Anc3*</th>
<th>Anc2*</th>
<th>Anc1*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AABUF ($\AA^2$)</td>
<td>Helicity (%)</td>
<td>AABUF ($\AA^2$)</td>
<td>Helicity (%)</td>
</tr>
<tr>
<td>Helix A</td>
<td>122.35</td>
<td>3.61</td>
<td>122.35</td>
<td>3.56</td>
</tr>
<tr>
<td>Helix D</td>
<td>135.97</td>
<td>8.55</td>
<td>130.04</td>
<td>5.11</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.90</td>
<td>0.42</td>
<td>0.94</td>
<td>0.70</td>
</tr>
<tr>
<td>Log2(Ratio)</td>
<td>-0.15</td>
<td>-1.24</td>
<td>-0.09</td>
<td>-0.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>AncA*</th>
<th>AncB*</th>
<th>AncC*</th>
<th>AncD*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AABUF ($\AA^2$)</td>
<td>Helicity (%)</td>
<td>AABUF ($\AA^2$)</td>
<td>Helicity (%)</td>
</tr>
<tr>
<td>Helix A</td>
<td>124.57</td>
<td>1.92</td>
<td>126.35</td>
<td>3.22</td>
</tr>
<tr>
<td>Helix D</td>
<td>130.36</td>
<td>2.59</td>
<td>131.49</td>
<td>4.63</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.96</td>
<td>0.74</td>
<td>0.96</td>
<td>0.69</td>
</tr>
<tr>
<td>Log2(Ratio)</td>
<td>-0.07</td>
<td>-0.43</td>
<td>-0.06</td>
<td>-0.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ecRNH*</th>
<th>ecRNH* A55G</th>
<th>ecRNH* D108L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AABUF ($\AA^2$)</td>
<td>Helicity (%)</td>
<td>AABUF ($\AA^2$)</td>
</tr>
<tr>
<td>Helix A</td>
<td>128.62</td>
<td>9.66</td>
<td>127.23</td>
</tr>
<tr>
<td>Helix D</td>
<td>129.95</td>
<td>2.47</td>
<td>129.95</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.99</td>
<td>3.92</td>
<td>0.98</td>
</tr>
<tr>
<td>Log2(Ratio)</td>
<td>-0.01</td>
<td>1.97</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

† AABUF values are the average across each helix as predicted using values from Rose, et al. (1985).
‡ Helicity values are the average across each helix as predicted using values from Muñoz and Serrano (1994).
Chapter 4:
Formation of an aggregation-prone folding Intermediate that is suppressed during translation
This article was published in Science Advances in 2018, and I am second author on this work. My contributions to this work include experimental, analytical, and interpretive work for the hydrogen exchange – mass spectrometry portions of the project.

4.1 Abstract

Precise protein folding is essential for the survival of all cells, and protein misfolding causes a number of diseases that lack effective therapies, yet the general principles governing protein folding in the cell remain poorly understood. In vivo, folding can begin cotranslationally and protein quality control at the ribosome is essential for cellular proteostasis. We directly characterize and compare the refolding and cotranslational folding trajectories of the protein HaloTag. We introduce new techniques for both measuring folding kinetics and detecting the conformations of partially folded intermediates during translation in real time. We find that, although translation does not affect the rate-limiting step of HaloTag folding, a key aggregation-prone intermediate observed during in vitro refolding experiments is no longer detectable. This rerouting of the folding pathway increases HaloTag’s folding efficiency and may serve as a general chaperone-independent mechanism of quality control by the ribosome.

4.2 Introduction

Biophysical characterization of protein energy landscapes has provided key insights into the mechanisms of protein folding and misfolding, design, and structure prediction. These in vitro studies, however, often fail to recapitulate the folding process in vivo. In the cell, the ribosome synthesizes proteins one amino acid at a time, and the translational machinery is a major hub for protein quality control. During translation, the nascent chain has the opportunity to explore regions of the energy landscape in the absence of the protein’s entire sequence. Therefore, cotranslational folding is fundamentally different from typical refolding experiments, where the full-length protein is denatured and then allowed to refold. Cotranslational folding has thus become a highly active area of research and has revealed insights into the mechanisms of protein-misfolding diseases.

Recent studies on stalled ribosome-nascent chain complexes (RNCs) have illuminated some of the features that guide cotranslational protein folding, that is, effects due to the tethering and proximity of the ribosome. These elegant studies, however, fail to recapitulate the dynamic process of translation; isolated, stalled RNCs are not sufficient for understanding the interplay between translation and protein folding. The importance of cotranslational folding is highlighted by biochemical studies, which have demonstrated that concurrent translation can increase the fidelity of protein folding and quaternary structure formation. Small changes in protein folding efficiency (the fraction of produced protein that folds to its native state) can have marked effects—they can overwhelm the cellular proteostasis machinery and lead to protein-misfolding diseases. Thus, proper cotranslational folding is essential for maintaining cellular and organismal proteostasis. The structural details for the folding process, however, have
only been characterized during in vitro refolding where the protein is refolded via dilution from a chemically or thermally denatured state. To understand how translation modulates protein folding, a direct comparison of the cotranslational and refolding trajectories is essential. However, we lack the high-resolution tools needed to monitor cotranslational folding.

Here, we determine the structural basis by which cotranslational folding increases the folding efficiency of the protein HaloTag and present new techniques to directly compare the structural and energetic differences between a protein’s cotranslational and standard refolding trajectories. We find that translation inhibits formation of a folding intermediate without changing the observed rate of overall folding, providing a general, chaperone-independent mechanism for increasing folding efficiency in vivo.

4.3 Results

**HaloTag refolding can be monitored by fluorescence polarization**

HaloTag is a modified haloalkane dehalogenase commonly used as a tool for in vivo imaging (Fig. 4-1, A and B) that covalently binds a ligand in its native conformation \[ k = 2.7 \times 10^6 \text{ M}^{-1} \text{s}^{-1}, \sim 27.0 \text{ s}^{-1} \text{ at } 10 \text{ μM tetramethylrhodamine (TMR)} \text{–ligand (the concentration used in this study)} \].\(^{24}\) If folding is much slower than 27 s\(^{-1}\), then we reasoned that the amount of protein-bound ligand in a folding experiment would be a direct measurement of folded HaloTag at any specific time. We monitored refolding of HaloTag in the presence of a free TMR-functionalized ligand using fluorescence polarization (FP). FP reports on the relative tumbling time of the fluorophore and thus is related to its apparent molecular weight. Rapid dilution of unfolded HaloTag into folding conditions (for example, 8 to 0.8 M urea), results in single exponential kinetics (\(k_{\text{obs}} = 4.7 \pm 0.9 \times 10^{-4} \text{ s}^{-1}\) at a final urea concentration of 0.8 M urea; Fig. 4-1). Linear extrapolation of the natural log of the folding rate, ln(\(k_{\text{obs}}\)), as a function of the final urea concentration\(^{25}\) yields a folding rate in the absence of a denaturant (\(k_{\text{H2O,FP}} = 4.8 \pm 0.6 \times 10^{-4} \text{ s}^{-1}\)) that is >10,000 times slower than ligand binding (Table 4-1) and similar to the folding rate determined by circular dichroism (CD; see below). Thus, changes in FP are measures of HaloTag folding.
Cotranslational folding can be monitored by fluorescence polarization

To monitor cotranslational folding in real time, we harnessed the same methodology, following FP, during in vitro translation. We initiated the IVT (in vitro transcription and translation) reaction directly in the fluorimeter by adding DNA encoding HaloTag to the coupled IVT system, PURExpress (New England Biolabs). Figure 4-1 reveals biphasic kinetics: a lag phase and an exponential phase. The observed kinetics are independent of the TM-ligand concentration and are specific to the HaloTag gene (Fig. 4-S1 and Table 4-S1). To confirm that the changes in TMR polarization monitor cotranslational folding and not protein synthesis, we independently determined the time
dependence of protein production using a gel-based assay (Fig. 4-1D). The observed kinetics of protein synthesis are also biphasic but with a lag phase significantly shorter than that observed by FP. In addition, we observed an exponential increase in FP signal even after the addition of the translation inhibitor neomycin, confirming that the change in FP reports on HaloTag folding and is not translation-limited (see Fig. 4-S1).

**Analysis of translation and folding kinetics**

These data were analyzed with a kinetic model to account for the asynchronous nature of both protein synthesis and protein folding (see Materials and Methods). The protein synthesis data were analyzed to determine the translation lag phase (251 ± 35 s), which represents the time to synthesize detectable protein levels, and the time-dependent translation rate. The average translation rate, ~1 aa s⁻¹, is similar to translation rates determined for other in vitro systems. To determine the

<table>
<thead>
<tr>
<th>Thermodynamic Data</th>
<th>$\Delta G_{CD}$ (kcal·mol⁻¹)</th>
<th>6.03±0.39</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m-value (kcal·mol⁻¹·M⁻¹)</td>
<td>1.57±0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kinetic Data</th>
<th>$\Delta G_{\text{kinetic}}$ (kcal·mol⁻¹)</th>
<th>5.24±2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m-value$_{\text{kinetic}}$ (kcal·mol⁻¹·M⁻¹)</td>
<td>1.41±0.58</td>
</tr>
<tr>
<td>$k_{H_2O}$ (sec⁻¹)</td>
<td>0.04±0.02</td>
<td></td>
</tr>
<tr>
<td>$m_{H_2O}$ (kcal·mol⁻¹·M⁻¹)</td>
<td>1.46±0.71</td>
<td></td>
</tr>
<tr>
<td>$k_{\text{constant, H}_2O}$ (sec⁻¹)</td>
<td>6.6±0.71x10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>$m_{\text{constant}}$ (kcal·mol⁻¹·M⁻¹)</td>
<td>0.02±0.1</td>
<td></td>
</tr>
<tr>
<td>$k_{H_2O, FP}$ (sec⁻¹)</td>
<td>4.8±0.6x10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>$m_{FP}$ (kcal·mol⁻¹·M⁻¹)</td>
<td>-0.1±0.04</td>
<td></td>
</tr>
<tr>
<td>$k_{NI, H_2O}$ (sec⁻¹)</td>
<td>8.47±20±10⁻⁶</td>
<td></td>
</tr>
<tr>
<td>$m_{NI}$ (kcal·mol⁻¹·M⁻¹)</td>
<td>-0.44±0.3</td>
<td></td>
</tr>
<tr>
<td>$k_{IU, H_2O}$ (sec⁻¹)</td>
<td>3.3±9.9x10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>$m_{IU}$ (kcal·mol⁻¹·M⁻¹)</td>
<td>0.70±0.15</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cysteine Accessibility</th>
<th>$k_{\text{WT, refolding, slow}}$ (sec⁻¹)</th>
<th>7.8±0.6±10⁻⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_{\text{WT, refolding, fast}}$ (sec⁻¹)</td>
<td>0.03±0.02</td>
</tr>
<tr>
<td></td>
<td>$k_{\text{WT, IVT}}$ (sec⁻¹)</td>
<td>4.7±0.3x10⁻⁴</td>
</tr>
<tr>
<td>$k_{M129C, \text{refolding}}$ (sec⁻¹)</td>
<td>&gt;0.01</td>
<td></td>
</tr>
<tr>
<td>$k_{M129C, \text{IVT}}$ (sec⁻¹)</td>
<td>3.2±0.2x10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>$k_{I126C, \text{refolding}}$ (sec⁻¹)</td>
<td>&gt;0.01</td>
<td></td>
</tr>
<tr>
<td>$k_{I126C, \text{IVT}}$ (sec⁻¹)</td>
<td>2.2±0.2x10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>$k_{E121C, \text{refolding}}$ (sec⁻¹)</td>
<td>&gt;0.01</td>
<td></td>
</tr>
<tr>
<td>$k_{E121C, \text{IVT}}$ (sec⁻¹)</td>
<td>&gt;0.01</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-1 Summary of kinetic and thermodynamic data of HaloTag and its mutants (Error bars are standard deviations of at least 3 separate measurements).
cotranslational folding rate, we augmented this model to include a protein-folding component. The lag time observed for the change in polarization is fourfold larger than that observed for protein synthesis, 811 ± 9 s versus 251 ± 35 s, respectively. The resulting cotranslational folding rate is thus 4.42 ± 0.02 × 10⁻₄ s⁻¹, similar to that obtained in the absence of translation (Table 4-1).

**Figure 4-2 Characterization of HaloTag folding kinetics and stability**

(A) Chevron plot of HaloTag folding and unfolding rates as a function of urea concentration. Fast phase (black circles) and slow phase (white circles, black outline). Refolding as measured by FP is shown in blue. Refolding traces of HaloTag at (B) 0.8 M urea, where there is visible protein aggregation, and (C) 1.6 M urea, where no precipitation is observed. (D) CD spectrum of HaloTag at 0 M urea. (E) Equilibrium denaturant melt of HaloTag. (F) Burst-phase amplitudes for refolding (white triangles with black outline) and unfolding (white squares with black outline). Kinetic final amplitudes (black circles) overlay well with the fit of equilibrium data (blue line). Error bars represent the SD of three separate experiments.
Refolding studies of HaloTag

We then characterized the stability and refolding of HaloTag using recombinant, purified protein (Fig. 4-2 and Table 4-1). Folding kinetics, determined by CD, fit to two exponential phases: a fast-folding phase and a urea-independent slow phase (Fig. 4-2, A to C). The slow phase, as measured by CD, corresponds to the refolding rate determined by FP (Table 4-1). Often, urea-independent folding is attributed to a cis-trans proline isomerization. However, both refolding and cotranslational folding in the presence of the proline isomerase cyclophilin A (CypA) revealed no effect, which suggests that this may not be due to proline isomerization (Fig. 4-S2, Tables 4-1, and Table 4-S3). Surprisingly, refolding to below 1.0 M urea resulted in visible precipitation and protein aggregation (Fig. 4-2 and Fig. 4-S3), although no aggregation was observed in the above cotranslational folding experiments that take place at 0 M urea. Aggregation occurred after an initial decrease in CD signal with a rate similar to the fast refolding phase observed in nonaggregating conditions. Using centrifugation, we determined the fraction of soluble protein to be 0.70 ± 0.06 under these conditions (Fig. 4-3A).

HaloTag cotranslational folding is more efficient than refolding

To compare the efficiencies of refolding and cotranslational folding, that is, the fraction of protein that reaches the native state, we used pulse proteolysis, a gel-based method for measuring the amount of folded protein (Fig. 4-3). Purified HaloTag incubated in 0.8 M urea is completely folded when evaluated by pulse proteolysis, but when refolded by dilution from 8.0 to 0.8 M urea, the efficiency is only 0.73 ± 0.03, consistent with that determined by centrifugation above. By contrast, cotranslational folding is significantly more efficient than refolding: 0.91 ± 0.03 versus 0.73 ± 0.03 (P < 0.01, Student’s unpaired t test; n > 12; Fig. 4-3 and Table 4-S2). Note that IVT reactions are carried out at a higher protein concentration than the less-efficient refolding studies (>5 and 3 μM, respectively; see Fig. 4-S1 and Materials and Methods). To rule out any possible chemical differences between in vitro–translated protein and recombinant protein, we measured the refolding efficiency of IVT protein and determined it to be similar to that of purified protein: 0.69 ± 0.06 versus 0.70 ± 0.06, respectively (Fig. 4-3). Why is cotranslational folding significantly more efficient than refolding? How does translation alter the folding pathway of HaloTag?

Structural characterization of the in vitro refolding pathway using HXMS

To compare the refolding and cotranslational folding pathways of HaloTag, we first used pulse-labeling hydrogen-deuterium exchange coupled with proteolysis and mass spectrometry (HX-MS) to obtain structural information about the conformational changes formed during HaloTag refolding. We applied pulses of hydrogen exchange at various refolding times and monitored the mass of individual peptides as a function of refolding time. Changes in mass are a measure of backbone amide accessibility at that particular refolding time. Figure 4-4A shows the fraction deuterated for each peptide
Figure 4-3 HaloTag folding is more efficient during in vitro translation than after refolding
(A) Fraction of total protein remaining in supernatant after centrifugation following refolding of HaloTag to 0.8 M urea. (B) Fraction folded as measured by pulse proteolysis in conditions as indicated—either after refolding, after in vitro translation, or both. Blue circles are in vitro-translated protein. (C) Representative gels for (A) and (B). All error bars are the SDs of at least 15 separate experiments except for HaloTag in 0.8 and 8.0 M urea, which are the SDs of three experiments. *P < 0.01, Student's unpaired t-test.
after 10 s of refolding and reveals two populations: those that are at least 25% deuterated by 10 s ("fast," red) and those that are not ("slow," blue). Plotting the mean-normalized fraction deuterated for both slow and fast peptides further highlights that these two groups of peptides have distinct behaviors throughout the folding trajectory (Fig. 4-4B). On average, fast peptides are more deuterated at all time points analyzed than slow peptides. Data for all peptides used in this analysis are available in Table 4-S4. These data, together with the biphasic CD kinetics (Fig. 4-2), suggest that the early protection is a result of the formation of a fast-folding intermediate. The early phase corresponds to protection of peptides comprising the Rossman-fold core of the protein, while the entire lid domain and β-strand 8 are protected more slowly (Fig. 4-4C). It is possible that the formation of this intermediate is directly responsible for HaloTag’s aggregation. For instance, helix B and β-strand 4 remain unprotected, despite the fact they both make critical contacts with the rest of the Rossman fold (β-strands 1 and 2 and helix C; Fig. 4-4C). This likely results in a large, exposed hydrophobic surface.

Thus, during the early steps of folding, not only is the lid domain completely unprotected but there is also a large exposed hydrophobic surface. Moreover, because this intermediate is comprised of residues distant in sequence space, it is possible that cotranslational folding does not involve formation of this intermediate.

Comparison of the HaloTag refolding and cotranslational folding trajectories using pulsed cysteine labeling

Unfortunately, the above HX-MS studies are currently not feasible for investigating cotranslational folding. Therefore, to determine whether HaloTag populates the same folding intermediate during cotranslational folding and refolding, we designed specific thiol probes based on the above HX-MS data. Labeling of reactive cysteines has been used successfully in the past on stalled RNCs. We monitored thiol accessibility during both refolding and cotranslational folding using a fluorescein-conjugated maleimide (FSM), detected by in-gel fluorescence.

Both of HaloTag’s native cysteines, positioned at the base of two β-strands, β4 and β8, are completely protected in the folded state and accessible in the unfolded state (modified within 30 s; see Fig. 4-S4). One of these, Cys262, is in the region we anticipate to be structured in the intermediate, and the other, Cys61, is not. Pulsed thiol labeling during refolding of the wild-type (WT) protein showed two phases with similar rates to those obtained by CD. By contrast, pulsed thiol labeling during cotranslational folding resulted in only a single exponential indistinguishable from HaloTag’s slow folding rate during refolding (Fig. 4-4D and Table 4-1).

We then created three site-specific cysteine variants to probe the very early stages of folding (E121C, I126C, and M129C) in an otherwise cysteine-free background (Halo*). Residues 126 and 129 are both buried side chains on β6, and during refolding, both are protected within the burst phase of the experiment (Fig. 4-4, E and F). E121C is on the surface of HaloTag and remains unprotected throughout the folding reaction (Fig. 4-4G). All three variants bind TMR and display similar folding kinetics as WT HaloTag (Fig. 4-S5). In contrast to the previous refolding experiments, during in vitro
Figure 4-4: The HaloTag folding trajectory changes during cotranslational folding
(A) Peptides derived from HX-MS experiments after 10 s of refolding were plotted according to their corresponding secondary structural element. Helices are lettered, whereas β sheets are numbered. Secondary structural elements were then divided into fast-folding (red circles) or slow-folding (blue circles) regions based on the average fraction deuterated (solid line) for peptides within those secondary structures at the 10-s time point (above or below dashed line). Error bars represent SEM. (B) Normalized fraction deuterated for all peptides (filled circles) plotted with the mean fraction deuterated for each group of peptides (solid lines) is shown for three time points. A full list of peptides is available in Table 4-S4. (C) Crystal structure of HaloTag with slow (blue) and fast (red) secondary structural elements colored. Loops are colored in white. Cysteines probed in (D) to (G) are represented as spheres (yellow, Cys$m^1$ and Cys$^m^2$; purple, M129C; blue, I126C; green, E121C). (D to G) Cysteine accessibility experiments during in vitro translation (colored lines and circles) and refolding (dotted lines and black dots). (D) WT HaloTag. (E) Halo* M129C. (F) Halo* I126C. (G) Halo* E121C. Error bars represent the SD of three separate experiments except for (A) where error bars are the SEM. Gels are shown in Figure 4-S6. AU, arbitrary units.

translation, sites 126 and 129 are not protected early but rather show slow protection
corresponding to the overall folding rate of the protein (Fig. 4-4D). Thus, the folding pathway of HaloTag is altered during translation.

4.4 Discussion

Together, our results suggest that the HaloTag refolding intermediate, which is likely the precursor for aggregation, is not populated during translation-coupled folding. This change in the folding trajectory is likely responsible for HaloTag's increased cotranslational folding efficiency. Moreover, this model also provides an explanation for the recent report that the mutation K73T, located within the structured region of the refolding intermediate, leads to increased HaloTag aggregation. The specific cysteines characterized here, however, do not yield further insight into other potential intermediates that may form during translation. The overall rate of folding is not changed during cotranslational folding, and thus, the rate-limiting step for folding does not appear to require the formation of this specific intermediate.

Intermediates in protein folding can play both positive and negative roles. Intermediates are often beneficial to the folding process by narrowing conformational space, while access to transient intermediates is also a major determinant for the formation of toxic aggregates associated with disease. Previous studies have suggested that formation of translation-specific intermediates may help to guide the folding process; our data support the hypothesis that destabilization of potentially toxic or off-pathway intermediates that form during translation are also advantageous. Thus, we have determined an additional mechanism by which translation helps to avoid aggregation of the emerging protein.

Our findings highlight the interplay between the rates of translation and folding. For instance, the relatively slow rate of translation in our IVT setup may aid in increasing HaloTag folding efficiency. Using the methods described here, it will now be possible to measure how folding efficiency and folding trajectories are modulated by the rate of translation.

HaloTag is ideally suited for these kinds of studies. HaloTag folding can be monitored by FP, thus folding experiments can be performed with high throughput and in the presence of many other biologically active molecules including during IVT. This is a powerful system to systematically investigate how the translational and quality control machinery modulates protein folding. These types of unbiased approaches will lead to the discovery of general and quantitative rules that govern not only protein folding during translation but also protein folding in other high-complexity environments.

4.5 Materials and Methods

Protein expression and purification

Protein expression. BL21(DE3) cells were transformed with expression vectors containing the WT or mutant HaloTag cDNA. Single colonies were used to seed starter cultures grown overnight to saturation. Large-scale cultures were inoculated with 5 ml of overnight culture, grown at 37°C to an optical density at 600 nm of 0.6 to 0.8 and
induced with 1 mM isopropyl-β-D-thiogalactopyranoside for 2 to 3 hours at 37°C. After induction, cultures were pelleted at 5000g for 10 min at 4°C, flash-frozen, and stored at −80°C.

Purification. Cell pellets were resuspended in 10 mM tris/H₂SO₄ (pH 7.5) and 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP; lysis buffer) and lysed by sonication on ice. Lysates were cleared by centrifugation for 30 min at 20,000g at 4°C and subsequently filtered through 0.2-μm filters. After clearing, the lysate was dialyzed into at least a 10-fold volume excess of lysis buffer, loaded onto a HiPrep Q XL 16/10 column equilibrated with lysis buffer, and eluted with a gradient of lysis buffer plus 0 to 600 mM NaCl. Fractions containing the HaloTag protein were dialyzed into at least a 10-fold volume excess of 20 mM sodium acetate (pH 5.0; Q buffer) loaded onto a HiPrep Q XL 16/10 column equilibrated with Q buffer and eluted with a gradient of Q buffer plus 0 to 800 mM NaCl. Fractions containing HaloTag protein were then concentrated and purified on a HiLoad 16/600 Superdex 75 pg column equilibrated with 50 mM ammonium bicarbonate or 25 mM Hepes KOH (pH 7.5), 15 mM MgOAc, 150 mM KCl, and 0.1 mM TCEP (HKMT), and the fractions with the retention volume corresponding to the size of monomeric HaloTag were either lyophilized (ammonium bicarbonate runs) and subsequently stored at −80°C or concentrated and immediately used for experiments (HKMT runs). All lyophilized protein was resuspended in HKMT and spun-filtered at 4°C before use in experiments.

**Fluorescence polarization**

Data collection. All experiments were performed at 37°C unless otherwise noted. FP was performed on a BioTek Synergy Neo2 plate reader in 384-well, black flat-bottom plates for IVT reactions (Corning) or 96-well, clear flat-bottom plates (refolding experiments). Acquisitions were collected using polarizers and 530-nm/590-nm filters with a side gain set at 45 and a top gain set at 40. Read height was 7.5 mm, and 10 measurements were made per data point. After a 15-min incubation at 37°C, readings were initialized by the addition of DNA (IVT reactions) or unfolded protein (refolding experiments). Measurements were taken every 20 s for 5 hours after 30 s of mixing and a 90-s delay for temperature equilibration.

IVT reactions. IVT reactions using a PURExpress system were set up on ice per the manufacturer’s protocols for a 30-μl reaction with the addition of 1 μl of ribonuclease (RNase) inhibitor, murine, and 1 μl of 300 μM TMR [in 100% anhydrous dimethyl sulfoxide (DMSO), for a final TMR concentration of 10 μM] and pipetted into wells. Plates were covered with clear titer tops to prevent evaporation and equilibrated at 37°C for at least 15 min. Reactions were initiated with 2 μl of plasmid DNA (125 ng/μl).

Refolding experiments. Refolding experiments were performed in HKMT buffer plus appropriate concentrations of urea and TMR (to a final concentration of 5 μM TMR and 3.33% DMSO). Plates were sealed and incubated at 37°C for 15 min until reactions were initiated by adding 10 μl of 20 μM HaloTag in 8 M urea that had been incubated at 37°C for at least 12 hours. Refolding traces were fit to the following equation in Matlab, using bi-square fitting and “k” bounded at zero.
\[ f(t) = a(1-e^{-kt})+c \]

Urea concentrations were measured using a refractometer as previously described.\textsuperscript{29}

**Circular dichroism**

Kinetic and equilibrium experiments were performed using a 0.5-cm cuvette at 37°C with constant stirring at 3 μM (0.1 mg/ml) in HKMT buffer. Equilibrium and kinetic experiments were performed as previously described\textsuperscript{40} but at a wavelength of 225 nm instead of 222 nm to increase the signal-to-noise ratio. Analysis was performed as described.\textsuperscript{40} Wavelength experiments were performed in a 0.1-cm cuvette at 37°C with 15 μM protein (~0.5 mg/ml) in HKMT buffer.

**Determination of folding efficiency**

All reactions were performed at 37°C at a final concentration of 3 μM protein in HKMT buffer unless otherwise noted.

Centrifugation assay. Proteins were refolded by the dilution of protein in 8 M urea to the proper urea concentration and allowed to reach equilibrium for at least 12 hours. Samples were then centrifuged at 21,130g for 30 min, and the supernatant was carefully removed. The pellet was resuspended in an equal volume of 8 M urea. Both the supernatant and pellet were then mixed with a 6× SDS–polyacrylamide gel electrophoresis (PAGE) loading dye and run on a 4 to 12% bis-tris gel in an MES run buffer and subsequently stained with SYPRO Red for 30 min in 10% acetic acid. After destaining in 10% acetic acid for an hour, gels were imaged using a Typhoon Trio (GE Healthcare) and analyzed with ImageJ.

Pulse proteolysis. IVT reactions were performed as per the manufacturer’s instructions but with the addition of 1 μl of RNase inhibitor, murine, and 1.25 μl of FluoroTect GreenLys (Promega) per 30 μl of IVT reaction (25). IVT reactions were quenched after 1 hour to a final concentration of 2 mM chloramphenicol and RNase A (0.1 mg/ml). Refolding experiments were performed as described above. IVT reactions and refolding reactions were allowed to reach equilibrium for at least 12 hours. Subsequently, reactions were aliquoted to 10 μl, and 1 μl of thermolysin (1 mg/ml; Sigma) was added to each reaction for 1 min and quenched with EDTA to a final concentration of 83 mM. SDS-PAGE loading dye was then added to each reaction, and each reaction was run on a 4 to 12% bis-tris gel in an MES run buffer. Imaging and analysis was performed as described previously.\textsuperscript{29}

Refolding of IVT-translated protein. IVT reactions were performed and quenched as described above. A 10-fold volume excess of 8 M urea in HKMT buffer was then added and mixed with the IVT translation reaction and allowed to equilibrate at 37°C overnight. Reactions were then concentrated in a 0.5-ml 10-kDa cutoff spin concentrator (Amicon) and diluted to 0.8 M urea. After equilibration at 37°C overnight, pulse proteolysis was performed and analyzed as described above.

**Translation rate measurement**
IVT reactions were performed as per the manufacturer’s instructions but with the addition of 1 μl of RNase inhibitor, murine, and 1.25 μl of FluoroTect GreenLys (Promega) per 30 μl of IVT reaction and initiated with 250 ng of DNA. At each time point, 1.5 μl of IVT reaction was quenched into a final concentration of 2 mM chloramphenicol and RNase A (0.1 mg/ml) and then SDS-PAGE loading dye. Reactions were then run on a 4 to 12% bis-tris gel in an MES run buffer and imaged using a Typhoon Trio. Analysis was performed using ImageJ.

**Cysteine protection assays**

Purified protein. Refolding reactions were initiated as described above. At each time point, a 50-fold molar excess of FSM was added for 30 s and quenched into an equal volume of SDS-PAGE loading dye containing β-mercaptoethanol (BME) to a final BME concentration of 2.15 M. Reactions were then run on 4 to 12% bis-tris gel in an MES run buffer and imaged using a Typhoon Trio. Analysis was performed using ImageJ. Traces were fit to the following equation in Matlab, using bi-square fitting and “k” bounded at zero (for those data which displayed exponential kinetics)

\[ f(t) = a(1 - e^{-kt}) + c \]

Because folded Halo**E121C cysteine reactivity is time-dependent over the labeling time of the reaction, intensities after refolding was initiated were normalized to the reactivity at that labeling time as determined in Fig. S7.

IVT reactions. IVT reactions were initiated as described above. At each time point, an equal volume of 2 mM FSM was mixed with IVT reaction for 30 s and quenched into SDS-PAGE loading dye as above. At 45 min, reactions were halted by the addition of chloramphenicol to a final concentration of 2 mM. Reactions were then run on a 4 to 12% bis-tris gel in an MES run buffer and imaged using a Typhoon Trio. Analysis was performed using ImageJ. Intensities were normalized to a major protein product running at ~65 kDa to control for effects of evaporation, fluorescein bleaching, and gel loading. Traces were fit to the following equation in Matlab after exclusion of points before 45 min, using bi-square fitting and “k” bounded at zero

\[ f(t) = a(1 - e^{-kt}) + c \]

**Pulse-Labeling HX-MS**

Pulse-labeling scheme. The HX-MS pulsed labeling experiments were based on previously described approaches.\(^{30,31}\) Deuterated protein was prepared by lyophilizing unfolded HaloTag in 8 M urea followed by resuspension in D\(_2\)O, repeated four times. Refolding experiments were carried out using a BioLogic QFM-4 apparatus in interrupted flow mode. To initiate refolding, we diluted deuterated protein in 8 M urea [deuterated HKMT buffer (pD\(_{read}\) 7.9), 10°C] with 10 volumes of deuterated HKMT to a final urea concentration of 1.6 M. After a variable delay time (refolding time), D-to-H
exchange at still-exposed sites was induced by a high-pH pulse of protonated buffer (200 mM glycine; 10 ms; 5 volumes; final pH, 10.00). Because of the large volume changes, the final solution is only 31% protonated. The pulse was quenched by dilution with a low-pH buffer (1 M glycine; 5 volumes; final pH, 2.00) to slow any further exchange. Protein samples were then collected and injected into a custom liquid chromatography–mass spectrometry (LC/MS) system. A folded control sample was prepared by subjecting deuterated, native protein to the same pulse/quench sequence, and an unfolded control was measured by performing the pulse/quench using fully deuterated, unfolded protein.

LC/MS system. A custom high-performance liquid chromatography system was used for in-line protease digestion, desalting, and separation of peptides. Peptides were eluted from the trap column and separated on an analytical C8 column using an acetonitrile gradient (5 to 90% acetonitrile) at 17 μl/min. The output of this system was directly injected into a Thermo Scientific LTQ Orbitrap Discovery using electrospray ionization.

Data analysis of HX-MS pulsed labeling. Peptides were identified using a SEQUEST search using a Proteome Discoverer 2.0 software. Peptide mass envelopes were fit using HDExaminer (Sierra Analytics) followed by a manual confirmation of each peptide. Deuterium content was assessed by examination of the centroid of each fitted peptide mass envelope. Only peptides with a high signal-to-noise ratio at each time point were used for further analysis. For each peptide at each time point, the fraction deuterated was determined by comparison to the folded and unfolded control samples.

**Kinetic modeling of translation and folding**

**Model for protein production in an IVT experiment**

To derive a chemical kinetic expression for the amount of protein produced as a function of time, we note that the rate of protein synthesis is the rate at which it takes for \(L\) amino acids to be covalently linked together to create a protein \(P\). Thus, the simplest reaction scheme for this process is

\[
L \rightarrow P. \tag{1}
\]

This reaction occurs with some rate, \(k\), which is a function of many processes including translation-initiation, -elongation and -termination. We consider \(L\) number of amino acids as one “bundle” (B) of monomer subunits; one bundle is required to synthesize one protein molecule. If \(N_{AA}\) is the total number of free amino acids in solution then the total number of bundles \(N_B\) is \(N_{AA}/L\). Thus, to express these reactants in concentration form we write the reaction scheme as

\[
B \rightarrow P, \tag{2}
\]

that is, state \(B\) consists of \(L\) free amino acids that are converted into \(P\).

The time evolution of the protein concentration, \([P(t)]\), is governed by the master equation

\[
\frac{d[P(t)]}{dt} = k[B(0) - P(t)]. \tag{3}
\]
Solving Eq. [3] under boundary conditions, \( P(t) = 0 \) at \( t = 0 \), yields
\[
[P(t)] = [B(0)](1 - e^{-kt}). \tag{4}
\]
The maximum protein concentration \( [P_{\text{max}}] \) at time \( t \to \infty \) is equal to \( [B(0)] \). Therefore,
\[
[P(t)] = [P_{\text{max}}](1 - e^{-kt}). \tag{5}
\]

There must be a time lag between the time at which transcription starts and the time at which the first protein molecule is fully synthesized. If this time lag is \( t_0 \) then Eq. [5] becomes
\[
[P(t)] = [P_{\text{max}}](1 - e^{-k(t-t_0)}). \tag{6}
\]

We fit the experimentally measured protein production curve using Eq. [6] and extracted the overall rate of protein synthesis \( k \) and time lag, \( t_0 \).

**Translation speed**

We estimated the average translation speed based on the method described in 41. We calculated the rate of protein production \( J(t) \) by taking the derivative of Eq. [6]
\[
J(t) = [P_{\text{max}}]ke^{-k(t-t_0)} \tag{7}
\]

Next, we normalized this quantity with the ribosome concentration (i.e., \( 0.45 \) \( \mu \)M), which we denote as \( j(t) \). Here \( j(t) \) is the rate of protein production from a single ribosome. On average then, \( 1/j(t) \) is the time required to synthesize a protein molecule and a ribosome goes through a round of initiation, elongation, termination and recycling during this time. Thus, \( 1/j(t) \) is an upper bound on the gene translation time, and if elongation is the rate limiting step then \( j(t)L \) is an estimate of the average codon translation rate.

**Analytical derivation for the fraction of folded protein**

To derive an expression for the probability of protein folding in an IVT experiment we assume that the protein is released in the unfolded state. Unfolded proteins fold post-translationally with rate \( k_F \) and the backward transition occurs with rate, \( k_U \). In this situation, the following chemical kinetic equations govern the time evolution of the concentration of proteins in the unfolded and folded state.

\[
\frac{d[U(t)]}{dt} = \frac{d[P(t)]}{dt} + k_U[F(t)] - k_F[U(t)] \tag{8}
\]

\[
\frac{d[F(t)]}{dt} = k_F[U(t)] - k_U[F(t)] \tag{9}
\]

and
\[
[U(t)] + [F(t)] = [P(t)]. \tag{10}
\]

\([F(t)] \) and \([U(t)] \) in Eqs. [8]-[10] are the concentration of proteins in the folded and unfolded state, respectively, whereas \([P(t)] \) (Eq. [6]) is the total protein...
concentration at time $t$.

Solving Eqs. [8]-[10] yields

$$\frac{[F(t)]}{[F_{\text{max}}]} = 1 + \frac{k}{k_F + k_U - k} e^{-k_F(t-t_0)} - \frac{k_F + k_U}{k_F + k_U - k} e^{-k(t-t_0)} \quad \text{when } t > t_0 \quad [11]$$

and

$$\frac{[F(t)]}{[F_{\text{max}}]} = 0 \quad \text{when } t \leq t_0$$

where $[F_{\text{max}}] = \frac{k_F [P_{\text{max}}]}{k_F + k_U}$ is the concentration of folded protein at $t \to \infty$. We used Eq. 11 to fit the experimentally measured folding probability $\left(\frac{[F(t)]}{[F_{\text{max}}]}\right)$ and extract the numerical value of $k_F$, assuming $k_U = 0$.

4.6 References


4.7 Supplemental Materials

Figure 4-S1. Cotranslational folding of HaloTag can be measured using FP
(A) Raw FP data (left axis) for IVT reactions initiated with HaloTag (black) and DHFR (gray) plasmids. Translation (right axis) of HaloTag as determined by gel. (B) Polarization as a function of TMR-ligand concentration during IVT of HaloTag. (C) Folding probability (left axis, blue lines) and HaloTag protein concentration (right axis, black dots with red line) as a function of time before and after the addition of neomycin. (D) and (E) Representative gels used to measure protein translation in figure S1A and figure S1C.
Figure 4-S2. Addition of the peptidyl-proline isomerase CypA does not affect HaloTag refolding or cotranslational folding rates

(A) Refolding of HaloTag in increasing concentrations of CypA as monitored by CD. (B) FP of HaloTag in the presence of 10μM CypA (blue) and no CypA (gray).
Figure 4-S3. Aggregation of HaloTag
HaloTag aggregates after refolding via dilution from 8.0M urea to the indicated final concentrations of urea.

Figure 4-S4. Cysteine accessibility of WT HaloTag
(A) Cysteine accessibility as a function of time as measured by fluorescein-maleimide fluorescence for unfolded (yellow circles) and folded (grey circles) HaloTag. (B) Raw data for plot in (A)
Figure 4-S5. Characterization of Halo* cysteine mutants

(A) Refolding rate as a function of [urea] as measured by FP for different HaloTag constructs. WT – yellow; Halo* M129C – purple; Halo* I126C – blue. (B) Cotranslational folding of HaloTag variants measured by FP. WT – yellow; Halo* M129C – purple; Halo* I126C – blue; Halo* E121C – green. (C) Cysteine accessibility as a function of time at 1.6M urea for Halo variants. (D) Cysteine accessibility as a fraction of unfolded intensity for refolded and native state Halo variants at 1.6M urea. (E) And (F) Gels used for plots in (C) and (D) respectively.
Figure 4-S6. Gels for data shown in Fig. 4
For (A)-(D), gel in color is during translation and gel shown in black and white is during refolding. * marks HaloTag bound to TMR-ligand for the purpose of finding the HaloTag band during analysis. (A) HaloTag WT. (B) Halo* M129C. (C) Halo* I126C. (D) Halo* E121C. (E) No template added.
Figure 4-S7. Characterization of Halo* E121C cysteine accessibility

(A) Cysteine accessibility of folded (black circles) and unfolded (white circles) Halo* E121C as measured by fluorescein-maleimide reactivity. (B) Gel used in (A)
<table>
<thead>
<tr>
<th></th>
<th>Lag Time (sec)</th>
<th>Rate ($10^{-4}$ sec$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refolding (polarization)</td>
<td>NA</td>
<td>4.8±0.6</td>
</tr>
<tr>
<td>Folding (<em>in vitro translation</em>)</td>
<td>811.29±8.57</td>
<td>4.42±0.02</td>
</tr>
<tr>
<td>Translation</td>
<td>251.29±34.6</td>
<td>4.50±0.10</td>
</tr>
<tr>
<td>Folding before +neo</td>
<td>724.18±17.0</td>
<td>5.49±0.07</td>
</tr>
<tr>
<td>Folding after +neo</td>
<td>NA</td>
<td>9.13±0.19</td>
</tr>
<tr>
<td>Folding[TMR] = 5uM</td>
<td>1065.89±17.32</td>
<td>4.39±0.04</td>
</tr>
<tr>
<td>Folding[TMR] = 10uM</td>
<td>811.29±8.6</td>
<td>4.42±0.02</td>
</tr>
<tr>
<td>Folding[TMR] = 12.5uM</td>
<td>780.05±8.9</td>
<td>6.13±0.04</td>
</tr>
</tbody>
</table>

Table 4-S1. Kinetic data obtained for HaloTag folding using FP

<table>
<thead>
<tr>
<th></th>
<th>Fraction Folded</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>1.04±0.01</td>
<td>3</td>
</tr>
<tr>
<td>Unfolded</td>
<td>0.0006±0.0003</td>
<td>3</td>
</tr>
<tr>
<td>0.8M refolded</td>
<td>0.73±0.10</td>
<td>15</td>
</tr>
<tr>
<td>0.8M dialysis</td>
<td>0.74±0.11</td>
<td>15</td>
</tr>
<tr>
<td>IVT refolded</td>
<td>0.69±0.06</td>
<td>15</td>
</tr>
<tr>
<td>IVT native</td>
<td>0.91±0.03</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 4-S2. Determination of HaloTag folding efficiency under different conditions

<table>
<thead>
<tr>
<th></th>
<th>Rate ($10^{-4}$ sec$^{-1}$)</th>
<th>$t_0$ (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (10μM TMR)</td>
<td>4.40±0.02</td>
<td>811.3±9</td>
</tr>
<tr>
<td>+10μM CypA</td>
<td>2.12±0.02</td>
<td>2007±70</td>
</tr>
<tr>
<td>Halo** I126C</td>
<td>2.08±0.02</td>
<td>612±33</td>
</tr>
<tr>
<td>Halo** M129C</td>
<td>2.20±0.08</td>
<td>848±10</td>
</tr>
<tr>
<td>Halo** E121C</td>
<td>2.32±0.11</td>
<td>755±57</td>
</tr>
</tbody>
</table>

Table 4-S3. Folding rates of HaloTag and variants measured by FP
<table>
<thead>
<tr>
<th>Time (s)</th>
<th>0</th>
<th>10</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folded</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S1/S2 Residues: 14-18</th>
<th>0.94984</th>
<th>0.847896</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1/S2 Residues: 15-18</td>
<td>0.988077</td>
<td>0.986077</td>
</tr>
<tr>
<td>S1/S2 Residues: 19-24</td>
<td>0.962235</td>
<td>0.956225</td>
</tr>
<tr>
<td>S1/S2 Residues: 24-36</td>
<td>1.062539</td>
<td>1.066259</td>
</tr>
<tr>
<td>S3 Residues: 37-42</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>S3 Residues: 37-45</td>
<td>0.080004</td>
<td>0.081038</td>
</tr>
<tr>
<td>S4 Residues: 37-45</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>A Residues: 37-54</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>A Residues: 47-54</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>A Residues: 47-54</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>A Residues: 47-54</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>B3/A Residues: 37-45</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>A Residues: 37-54</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>A Residues: 47-54</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>A Residues: 47-54</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
<tr>
<td>S5/C Residues: 103-117</td>
<td>0.039574</td>
<td>0.041534</td>
</tr>
</tbody>
</table>

Normalized HX-MS peptides used in Figure 4 Peptides in blue were excluded from analysis.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.69259</td>
<td>0.69466</td>
<td>0.70401</td>
<td>0.70440</td>
<td>0.70979</td>
<td>0.71936</td>
<td>0.72057</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
<td>0.72165</td>
</tr>
<tr>
<td>0.1</td>
<td>0.70794</td>
<td>0.70838</td>
<td>0.71481</td>
<td>0.71525</td>
<td>0.71706</td>
<td>0.71883</td>
<td>0.72044</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
<td>0.72111</td>
</tr>
<tr>
<td>0.2</td>
<td>0.71237</td>
<td>0.71281</td>
<td>0.71854</td>
<td>0.71898</td>
<td>0.72083</td>
<td>0.72257</td>
<td>0.72419</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
<td>0.72486</td>
</tr>
<tr>
<td>0.3</td>
<td>0.71681</td>
<td>0.71725</td>
<td>0.72349</td>
<td>0.72393</td>
<td>0.72579</td>
<td>0.72754</td>
<td>0.72918</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
<td>0.72985</td>
</tr>
<tr>
<td>0.4</td>
<td>0.72127</td>
<td>0.72171</td>
<td>0.72793</td>
<td>0.72837</td>
<td>0.72992</td>
<td>0.73166</td>
<td>0.73331</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
<td>0.73398</td>
</tr>
<tr>
<td>0.5</td>
<td>0.72571</td>
<td>0.72615</td>
<td>0.73231</td>
<td>0.73275</td>
<td>0.73441</td>
<td>0.73615</td>
<td>0.73780</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
<td>0.73846</td>
</tr>
<tr>
<td>0.6</td>
<td>0.72993</td>
<td>0.73038</td>
<td>0.73552</td>
<td>0.73597</td>
<td>0.73763</td>
<td>0.73937</td>
<td>0.74102</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
<td>0.74168</td>
</tr>
<tr>
<td>0.7</td>
<td>0.73421</td>
<td>0.73465</td>
<td>0.73975</td>
<td>0.73920</td>
<td>0.74086</td>
<td>0.74260</td>
<td>0.74425</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
<td>0.74491</td>
</tr>
<tr>
<td>0.8</td>
<td>0.73848</td>
<td>0.73893</td>
<td>0.74408</td>
<td>0.74353</td>
<td>0.74519</td>
<td>0.74693</td>
<td>0.74858</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
<td>0.74924</td>
</tr>
<tr>
<td>0.9</td>
<td>0.74275</td>
<td>0.74320</td>
<td>0.74836</td>
<td>0.74781</td>
<td>0.74947</td>
<td>0.75121</td>
<td>0.75286</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
<td>0.75352</td>
</tr>
<tr>
<td>1.0</td>
<td>0.74702</td>
<td>0.74747</td>
<td>0.75215</td>
<td>0.75160</td>
<td>0.75326</td>
<td>0.75500</td>
<td>0.75665</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
<td>0.75731</td>
</tr>
</tbody>
</table>

Folded
I Residues: 228-234
Loop Residues: 234-239
S7 Residues: 238-243
S7 Residues: 240-246
J Residues: 244-255
S8 Residues: 263-273
S8 Residues: 265-273
S8 Residues: 267-273
K/L Residues: 274-285
L Residues: 286-292
S4 Residues: 61-68
B Residues: 81-86
B Residues: 91-95
B Residues: 98-103

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Folded</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residues</th>
<th>Time (s)</th>
<th>Folded</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.410477</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.417467</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.39854</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.550477</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.55775</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.58033</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.61589</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.66539</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.68694</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.71771</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.74923</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.77761</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.80758</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.84425</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.87126</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.90772</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.94317</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.97966</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.01631</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.05378</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.09125</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.12872</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.16619</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.20367</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.24114</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.27872</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.31630</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.35386</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.39143</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.42900</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.46656</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.50413</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.54170</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.57927</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.61683</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.65440</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.69196</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.72953</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.76710</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.80467</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.84223</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.87980</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.91737</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.95493</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.99250</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Chapter 5:
Increased sequence resolution of hydrogen exchange mass spectrometry studies through sub-peptide level protection factors
5.1 Abstract
Advances in the use of mass spectrometry as a detection method for hydrogen exchange experiments have allowed massive quantities of data on protein structure and stability to be generated quickly. Despite these advances, structural interpretation of these data is limited to relative comparison between different conditions and few methods quantitating protection in MS studies exist. Here, we expand on previous work by introducing a method to identify protection factors at the sub-peptide level from hydrogen exchange-mass spectrometry (HX-MS) data. This method, based on an empirical comparison of label uptake curves, is capable of generating multiple protection factors across a single peptide within an order of magnitude. Importantly, this method correctly identifies these protection factors more frequently than current methods. This method is robust across a wide range of protection factor distributions, and the protection factors extracted using this method are consistent with residue specific protection factors determined using NMR on the model protein RNase H from *E. coli*.

5.2 Introduction
The ability to detect protein thermodynamics and kinetics in a non-perturbing environment has made hydrogen exchange a premier technique for examining many aspects of protein structure, stability, and interaction. The quantitative interpretation of hydrogen exchange data is most informative when used to identify per-residue protection factors (PFs) which are a measure of how much the exchange of amide hydrogens at each peptide bond is slowed by protein structure. Because detection of exchange by NMR depends on the assignments of individual amide protons within the NMR spectrum, NMR detection can clearly assign a protection factor to each observed residue allowing exquisite structural resolution on protein structure, dynamics, and stability.\(^1,2\) Bottom-up mass spectrometry detection can overcome many inherent limitations present in NMR, but this method presents the added complication that multiple amide exchanges are detected within each peptide, and experimental limitations often prevent assigning individual exchange rates to individual residues reducing this sequence resolution. Thus typical hydrogen exchange mass spectrometry (HX-MS) workflows are limited to a comparison of individual peptides under different conditions preventing quantitative comparison with other experimental work.

To address the heterogeneity in exchange rates in a single peptide, recent attempts at quantitative interpretation of HX-MS data fit peptide exchange to multiple or stretched exponentials or use global fitting to identify individual protection factors.\(^3-6\) Overlapping peptides and high resolution data have allowed determination of individual exchange rates under certain conditions.\(^3,7,8\) One study used a combination of HX with NMR detection and HX-MS to identify exchange behavior on the sub-peptide level.\(^9\) Each of these approaches can identify PFs under certain, strict conditions. Recent work, however, has demonstrated a general approach for identification of peptide-level average protection factors using a model-free treatment of the data.\(^10,11\) This recent approach, which involves quantifying the slowing of exchange at the peptide level either in terms of a geometric approximation of the area between uptake curves or as the distance between these curves at multiple points, allows one to compare the exchange process of a peptide under multiple conditions. While extremely powerful, these
approaches are still limited by experimental constraints that often capture only a subset of the exchanges across each peptide. In addition, they produce only peptide-averaged protection factors, which have limited utility when a small number of long peptides are observed as is common in complex systems.

In order to identify multiple protection factors for a peptide, we have harnessed the information in the distribution of predicted protection factors derived from model-free methods. Combining this with overlapping peptides generated from nonspecific proteolysis allows quantitative description of protection factors which are resolved at a sub-peptide level. To test the utility of this method, we examine the protection factors of the Ribonuclease H protein from *E. coli*. This protein is known to occupy several partially unfolded forms (PUFs) at equilibrium under native conditions, which generate a spectrum of protection factors under native conditions that have previously been determined by HX-NMR methods. In addition, recent studies have optimized the HX-MS workflow for this protein.\(^1,12\)

5.3 Method

Protection factors are a critical metric to quantify the results of hydrogen exchange for comparison across multiple experiments. In the most basic sense, protection factors are merely the ratio of the intrinsic rate of exchange to the observed rate of exchange at a given amide site \(k_{\text{int}}/k_{\text{obs}}\). However, the quantification of protection factors is difficult when measurement is performed by mass spectrometry as the peptides detected in MS experiments have multiple exchangeable amides and it is

---

**Figure 5-1 Analysis of Peptide Deuteration**

(A) Deuteration level for a peptide at different exchange times, herein called samples, (blue dots) are measured using Mass Spectrometry. (B) Fitting the uptake of deuterons to the sum of exponentials (dotted line) gives individual exchange rates. (C) Using linear interpolation to connect the measured points (solid blue lines) and then the uptake curve of the intrinsic rate of exchange for the peptide (black line). By taking the ratio of x-axis values (TR values) in the blue and black lines at points evenly-spaced along the y-axis (black dotted lines), we can extract the average protection factor, \(<PF>\), across the whole peptide. (D) By taking the distribution of TR values and fitting to the sum of skew normal distributions (red line), we can extract multiple PF values per peptide.
often impossible to distinguish exchange at one site from that at another. The two predominant methods of identifying protection factors (PFs) in MS data are outlined in Figure 5-1. In the first method, isotopic labeling of peptides as a function of time (represented on a log scale) can be fit directly to the sum of exponentials to obtain observed rate(s) of exchange, \( k_{\text{exp}} \) (see Supplemental Methods) (Figure 5-1A-B). Observed rate(s) can then be compared to the intrinsic rate of exchange for an amide site, \( k_{\text{int}} \), or the average intrinsic rate of multiple sites to find a protection factor. In the second method the entire uptake curve of label versus time for a peptide is compared to the uptake curve in the absence of structure (sum of all intrinsic rates in the peptide) to empirically determine the average protection factor at the peptide level (Figure 5-1C). In this method the labeling time that is required for a peptide to exchange a given number of backbone sites labeled (\( d \)) under experimental conditions (\( t_{\text{exp}} \)) is compared to the time it would take if all amides in that peptide exchanged with their intrinsic rate (\( t_{\text{int}} \)). The ratio \( t_{\text{exp}} / t_{\text{int}} \), herein called TR, is a measure of the protection, and for the case of a single exchangeable amide is constant for all values of \( d \) and approximates the protection factor of that amide. For peptides with multiple exchange sites and multiple protection factors, Walters showed that the geometric mean of these TR ratios at multiple evenly distributed values of \( d \) is approximately equivalent to the average protection factor for these sites (\(<PF>\)).

The distribution of these TR values contains further information on the number of different protection factors in a peptide and their approximate values. By fitting the distribution of the TR values to the sum of multiple skew normal distributions (see Supplemental Information), we can find the number of summed distributions (number of protection factors) and their respective centroids (approximate value) present in a single peptide (Figure 5-1D). This allows us to empirically determine protection factors at a sub-peptide level from peptide-level data without the need to directly fit amide exchange rates. In addition to identifying these protection factors, we can use overlapping peptides to isolate these protection factors in the sequence at a sub-peptide level, similar to previous work.\(^7\,\,13\)

![Figure 5-2 Error in fitting PFs](image)

100 simulated peptides were fit to determine the accuracy of the method in determining the protection factor. The RMSD in the Log\(_{10}(PF)\) value predicted by fitting the TR values as a function of different parameters in the fitting are shown. N=100 at each point for each protection factor. Total values are the RMSD for all protection factors combined. (A) Error as a function of the number of samples (x/y points) taken when making the deuterium uptake curve. (B) Error as a function of the number of comparison points taken along the y-axis (TR values). Subsequent simulations used 15 samples and 200 TR comparisons.
HX-MS data for the exchange of a cysteine-free variant of the RNase H protein from *E. coli*, ecRNH*, were obtained for comparison to NMR work. Exchange was carried out at pH 5.5 over the course of 6 months in deuterated buffer with varying concentrations of denaturant, and aliquots were quenched, flash frozen, and stored at -80°C. Detection using mass spectrometry was carried out as described in Supplemental Information. Of >130 peptides, 36 were observed in all timepoints and denaturant concentrations and fitting of the TR distribution of these peptides was used to predict protection factors for each peptide (Supplemental Information).

5.4 Results and Discussion

Robustness of Method

To first test the robustness of fitting the TR distribution to obtain protection factors, simulated peptide exchange curves were used to generate TR distributions that were then fit to one or the sum of two skew normal distributions and the accuracy of the predicted PFs was examined. This accuracy was measured as a function of the number of timepoints taken, herein called samples, and the number of TR comparisons along the curve that were used, herein called comparisons (Supplemental Information) (Figures 5-2). One hundred simulated peptides each representing ten exchangeable amides with a uniform protection factor throughout were used to determine the accuracy of the predicted PF value at a range of sample and comparison numbers. Unsurprisingly, as the number of samples increases, the predicted PF from fitting the TR distribution becomes closer to the actual value. This remained true over a wide range of PF values (10\(^{-1}\)-10\(^8\)). In addition, increasing the number comparison points increased accuracy. The accuracy for this method reached a plateau around 15 samples and 100 comparisons. For comparison to existing methods, a similar analysis was performed by fitting simulated exchange curves to the sum of exponentials to find rates of exchange, and both of these analyses were repeated with simulated peptides containing multiple protection factors. These showed similar behavior with increased accuracy as sample and comparison number increased and accuracy plateaued at similar values (Supplemental Information, Supplemental Figures 5-S1, 5-S2).

To further test the accuracy of each method, both fitting of rates and fitting of TR distributions were performed on 1000 simulated peptides at a range of protection factor values and combinations of protection factors. While fitting the rate of exchange was extremely accurate when only a single protection factor is used to generate a peptide, fitting the TR values results in an estimate of the protection factor that is within one order of magnitude (Figure 5-3A). When two separate protection factors were used to simulate peptides, both methods were roughly equally accurate at predicting PF values with fitting of rates having a slight edge over fitting TRs (Figure 5-3B).

While the accuracy of the PF value did not change across a wide range of protection factors, the ability for each method to accurately determine the presence of multiple protection factors varied greatly. Both methods were better able to determine the presence of multiple PFs if the PF values were very different, but the method of fitting TR distribution was better able to correctly discern multiple distinct protection factors similar in value beginning when the underlying PFs were separated by one order of magnitude (Figure 5-3C, Supplemental Figure 5-S3). In contrast, directly fitting the rate of uptake gives an equally good fit to a single exponential unless the underlying protection factors are separated by at least two orders of magnitude. While many
previous studies have identified large differences in protection across a protein, as well as within a single peptide, the increased sensitivity of this method can allow detection of subtle differences in stability across a peptide.

Native-state Hydrogen Exchange of RNase H

To test this method on a known system, TR distribution fitting was then used to extract multiple protection factors per peptide from a dataset examining the native state of the ecRNH* protein. Continuous labeling hydrogen exchange was carried out at pH 5.5 over 6 months in various concentrations of denaturant. Previous work using NMR detection has shown that this protein populates two partially-folded forms under these conditions. Each HX-MS experiment produced >130 peptides that were detected at every timepoint. Of these, 36 peptides were seen in every timepoint in all denaturant concentrations and were used for further analysis. Of these peptides, previous work identifying PFs at residue resolution using NMR suggests at least 26 (72%) contain multiple protection factors that differ by more than one order of magnitude highlighting the need for moving away from using peptide-averaged protection factors. Using the method of fitting the TR distribution, 22 peptides (61%) were fit to have multiple distinct protection factors. In addition, many peptides had residues that exchanged faster or slower than the limit of detection. In this experiment these limits correspond to Figure 5-3 Comparison of TR fitting and exponential fitting

Simulated peptides were fit to determine the accuracy of multiple methods in determining the protection factor from HX-MS data. Fitting of the TR distribution is shown in blue and fitting the rate of exchange to one or multiple exponentials are shown in red. Standard deviations are shown as filled area around the curves. N=1000 for each point. Dotted lines represent the value of protection factors used to generate peptides. (A) The performance of both methods at a range of protection factors when a single protection factor is used to simulate protection factors. (B) Same as (A), but two protection factors are used to simulate each peptide (PF1=10, PF2=10^2-10^8). (C) Peptides were generated with two protection factors (PF1=10, PF2=10^2-10^8) and the number of peptides each method correctly fit as two protection factors is shown for each PF distribution.
Protection factors of ~5 and ~$10^5$. This means that for each peptide, we now get information on the number of very weakly (PF<5) and very strongly (PF>$10^5$) protected amide sites as well as the number of residues and amount of protection for intermediately protected (1 or 2 PFs where 5<PF<$10^5$) amides.

To compare the residue-level protection factors obtained using this method, we directly compared the protection factors extracted from TR distribution fitting to protection factors extracted from NMR. Despite the increased resolution afforded by this method, the limited number of peptides used in this analysis resulted in very few residues where overlapping peptides can be used to confidently assign a protection factor to an exact residue. The one residue that is site resolved in both the NMR and MS experiments, K86, shows similar protection factors in both experiments ($\Delta G_{\text{HX-NMR}} = 5.66$ kcal/mol, $\Delta G_{\text{HX-MS}} = 5.64$ kcal/mol) (Supplemental Info). Despite the small number of site resolved protection factors in the mass spec experiment, it is possible to discern the average PF for a small (sub-peptide level) group of residues. To further compare the MS and NMR datasets, these averages for residues were compared to the site-resolved
PFs from NMR. This comparison is shown in Figure 5-4. The data show a weak correlation between site-resolved NMR and sub-peptide-averaged MS protection factors with an R² of 0.296. Despite the inability to achieve single-residue resolution, mass spectrometry detection allows the most dynamic regions of a protein to be probed and average protection across this region to be quantitated. This work was able to detect and assign average protection to the loop region of ecRNH at ΔG_HX ~ 2.4 kcal/mol, well below the lowest detected exchanges in the NMR experiment (ΔG_HX ~ 3.5 kcal/mol). This presents an exciting possibility to examine previously inaccessible proteins or protein regions using HX-MS. These data suggest that while mass spec detection can give increased coverage and TR distribution fitting can give better resolution, achieving residue-level resolution that rivals NMR work will require large numbers of overlapping peptides and cannot rely on peptide- or sub-peptide-level averaged PFs.

5.6 Conclusions
The method of using time ratios (TR values) of deuterium uptake curves to empirically identify protection factors is a powerful tool that can improve upon current quantitative methods. Analyzing the distribution of TR values for a peptide can give an estimate of multiple protection factors per peptide increasing sequence resolution of protection. Using this method, mass spectrometry studies can be used to examine proteins with sufficient detail to allow comparison to NMR studies.

5.7 References


### 5.7 Supplemental Information

*Simulating exchange curves*

Peptide exchange curves were simulated using intrinsic rates \((k_{int})\) from 10 consecutive residues of the RNase H protein from *E. coli* at pH 5.5. Protection factors were assigned to each of these sites randomly and used to find an exchange rate \((k_{ex}=k_{int}/PF)\) for each site. Exchange curves were then calculated by calculating the sum of these exponentials at different x-values.
\[ y = \sum (1 - e^{-k_{ex,i}t}) \]

These x-values were evenly spaced along a logarithmic scale and the number of x-values was varied to determine robustness and then fixed at 15. Noise was then added to each point in both x- (0.2 orders of magnitude) and y- (0.4 deuterons) directions to result in the final curve. Peptides were simulated with protection factors ranging from 10\(^{-10}\) to 10\(^{8}\) and for peptides with multiple protection factors, 5 residues were given PF=10 and 5 residues were given PF=10\(^{2}\) to 10\(^{8}\).

**Method for determining \(<PF>\) and TR distribution**

Described in detail in previous work,\(^1\) briefly deuterium uptake for a peptide is plotted on a semilog plot characterizing the number of deuterons exchanged versus log(time). A linear interpolation is used to connect these points, and then evenly spaced values along the y-axis are used for comparison. For each of these y-values, the x-value on the interpolated curve \(t_{exp}\) is compared to the intrinsic exchange curve at the same y-value \(t_{int}\). The ratio of these two values \(TR = t_{exp}/t_{int}\) gives a measure of protection, and the geometric mean of all TR values gives the approximate average protection of all sites in the peptide \(<PF>\). The number of TR values taken per peptide was varied to determine robustness and then fixed at 200.

**Fitting of TR distribution**

To determine the underlying protection factor, comparisons (TR values) along the entire deuterium uptake curve are taken and the distribution is examined. The TR value distribution is fit to either one or the sum of two skew normal distributions (Equations 1 and 2) with amplitude \(A\), skew \(\alpha\), location \(\xi\), and scale \(\omega\).

\[
A_2 \frac{2}{\omega_2 \sqrt{2\pi}} e^{-\frac{(x-\xi)^2}{2\omega_2^2}} \int_{-\infty}^{\frac{x-\xi}{\omega_2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt 
\]

\[
A_1 \frac{2}{\omega_1 \sqrt{2\pi}} e^{-\frac{(x-\xi_1)^2}{2\omega_1^2}} \int_{-\infty}^{\frac{x-\xi_1}{\omega_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + A_2 \frac{2}{\omega_2 \sqrt{2\pi}} e^{-\frac{(x-\xi_2)^2}{2\omega_2^2}} \int_{-\infty}^{\frac{x-\xi_2}{\omega_2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt 
\]  (Eq. 2)

The locations \((\xi)\) of each distribution can be used as the average TR for that distribution. If one of the uptake curves used for this comparison is the intrinsic rate of exchange, this TR value \(TR = t_{exp}/t_{int}\) is the protection factor \(PF\) for the amides giving rise to the distribution.

**Fitting of exchange curve to the sum of exponentials**

Direct fitting of the uptake curve to the sum of one or two exponentials was performed for comparison to the TR distribution method. The amount of deuterium (y) versus time is fit to equations 3 and 4.

\[ y = A \cdot e^{-k_{exp}t} \]  (Eq. 3)

\[ y = A_1 \cdot e^{-k_{exp,1}t} + A_2 \cdot e^{-k_{exp,2}t} \]  (Eq. 4)

In order to obtain protection factors from these fits the fitted rate must be compared to an intrinsic exchange rate. The rate used for comparison is the geometric mean of the intrinsic exchange rates of all measurable amides in the peptide (Equation 5).

\[
k_{int,avr} = \sqrt[N]{\prod_{i=1}^{N} k_{int,i}} \]  (Eq. 5)

Where \(N\) = number of exchangeable amides in the peptide. It is assumed that the first two amides for each peptide will back exchange during the workup and measurement by mass spectrometry, so these rates are ignored in the calculation of the average rate.
The protection factors are then calculated by dividing the intrinsic rate of exchange by the observed rate of exchange (Equation 6).

\[ PF = \frac{k_{\text{exp}}}{<k_{\text{int}}>} \quad \text{(Eq. 6)} \]

**Data Analysis for Simulated Peptides**

When determining the minimum number of x-values or “samples” per peptide required for robust determination of the PF, the RMSD for the predicted PF was calculated for 100 peptides simulated per sample number (Figure 2A). A similar process was used to determine the number of TR values used (Figure 2B). This process was repeated using fitting to the sum of exponentials (Figure 5-S1). Once these values were fixed, 1000 peptides were simulated with each PF distribution with sampling at 15 x-values and 200 TR comparisons and the predicted PF values from these are plotted in Figure 2C-D with their standard deviation.

**Hydrogen Exchange of ecRNH**

The cysteine-free version of the RNase H protein from E. coli was expressed and purified as described previously.\(^4\) Protein stock was made at 160 μM in 100 mM NaOAc at pH 5.5 with 0, 1, and 1.2M GdmCl. Protein stock was diluted 1:9 in deuterated in 100 mM NaOAc at pD\(_{\text{corr}}\) = 5.5. At various labeling times, aliquots of reaction were taken and quenched with 1:3 dilution in 500 mM Glycine pH 2.2, 1M GdmCl and then flash frozen and stored at -80°C until mass spec detection. Timepoints at 0M GdmCl were taken over 6 months; in 1 and 1.2M GdmCl timepoints were taken over 14 days.

**Mass Spectrometry Detection**

Detection of peptide deuteration levels was carried out using a Thermo Orbitrap Discovery mass spectrometer. Quenched deuterated protein samples were thawed on ice and injected into a custom HPLC system for protein digestion and peptide separation. Protein digestion was carried out using on-line immobilized pepsin and fungal protease XIII in 0.1% formic acid. Peptide desalting and separation were carried out using C4 and C8 columns and a 0-90% acetonitrile gradient in 0.1% formic acid. Ionization was performed using a Thermo IonMax ESI source. Peptide identification was carried out using tandem mass spectrometry of undeuterated samples and Proteome Discoverer 2.0 Software, and deuteration levels were obtained with HDExaminer (Sierra Analytics). Centroids of peptide masses were used to determine deuterium uptake.

**Data Analysis for NSHX on ecRNH**

Peptide deuteration levels were taken from the program HDExaminer, and only peptides that were present in all timepoints and at all denaturant concentrations were used for further analysis. TR value determination and distribution fitting were carried out as described above to extract (a) PF value(s) for each peptide. Then, the TR value versus deuterons exchanged graph was examined for each peptide in order to quantify the number of residues exchanging with each protection factor (Supplemental Figure 5-S4). Overlapping peptides were then used to isolate which residues contain each protection factor to the smallest sequence window possible. Finally, PF values for each residue were refined by taking the average of the PF values from all peptides that report on the PF in that window. For calculating free energies of exchange (\(\Delta G_{\text{HX}}\)) the protection factor, P, is used in the equation \(\Delta G_{\text{HX}} = -RT \ln(P)\).
Supplemental Figures

Figure 5-S1 Rate fitting accuracy
The RMSD for protection factors predicted from direct fitting of the rate of exchange as a function of the number of samples. Peptides were generated with a single protection factor at a range of protection factors. N=100 for each point along each colored line. Black line represents the total RMSD for all protection factors at each sample number.
Figure 5-S2 Accuracy of fitting of multiple PFs per peptide
A-B) Accuracy of fitting the rate of exchange for simulated peptides with multiple protection factors across a range of sample numbers. A) represents the accuracy of the smaller protection factor, B) represents the accuracy of the larger protection factor. C-F) Accuracy of fitting the TR distribution for simulated peptides across a range of sample numbers (C & D) and number of TR comparisons (E & F). C) & E) represent the accuracy of the smaller protection factor, D) & F) represent the accuracy of the larger protection factor.
Figure 5-S3 Single Protection Factor Fitting
The fraction of peptides that were correctly fit with a single protection factor by each method. N=1000 for each bar. Peptides were simulated using a single protection factor (PF=10^1-10^8) and 15 sample points. TR fitting used 200 comparison points for TR distribution fitting.

Figure 5-S4 Plot of TR values
Example plot for the TR values for a peptide derived from taking $\frac{t_{exp}}{t_{int}}$ evenly spaced along the deuteron exchange events observed. This graph can be used to determine the number residues exchanging with each protection factor.
<table>
<thead>
<tr>
<th>Residue</th>
<th>$\Delta G_{0M}$ (kcal/mol)</th>
<th>$\Delta G_{1M}$ (kcal/mol)</th>
<th>$\Delta G_{1.2M}$ (kcal/mol)</th>
<th>Resolved region</th>
<th>0M Fit (y-intercept)</th>
<th>m-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.40</td>
<td>2.88</td>
<td>2.56</td>
<td>1</td>
<td>4.51</td>
<td>-1.63</td>
</tr>
<tr>
<td>4</td>
<td>5.40</td>
<td>2.88</td>
<td>2.56</td>
<td>1</td>
<td>4.51</td>
<td>-1.63</td>
</tr>
<tr>
<td>5</td>
<td>4.65</td>
<td>0.75</td>
<td>&lt;0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4.65</td>
<td>0.75</td>
<td>&lt;0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6.21</td>
<td>2.88</td>
<td>2.56</td>
<td>2</td>
<td>4.51</td>
<td>-1.63</td>
</tr>
<tr>
<td>8</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>9</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>10</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>11</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>12</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>13</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>14</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>15</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>16</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>17</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>20</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>22</td>
<td>4.76</td>
<td>3.04</td>
<td>2.56</td>
<td>3</td>
<td>5.44</td>
<td>-2.40</td>
</tr>
<tr>
<td>23</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>28</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>29</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>30</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>31</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>32</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>33</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>34</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>35</td>
<td>6.3338206</td>
<td>3.09</td>
<td>2.46</td>
<td>4</td>
<td>6.21</td>
<td>-3.12</td>
</tr>
<tr>
<td>36</td>
<td>6.3338206</td>
<td>&lt;3</td>
<td>&lt;3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>42</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>43</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>45</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>46</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>47</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>48</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>49</td>
<td>4.34</td>
<td>3.81</td>
<td>3.36</td>
<td>5</td>
<td>6.05</td>
<td>-2.25</td>
</tr>
<tr>
<td>50</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>7.05</td>
<td>4.13</td>
<td>2.90</td>
<td>7</td>
<td>10.30</td>
<td>-6.17</td>
</tr>
<tr>
<td>57</td>
<td>7.05</td>
<td>4.13</td>
<td>2.90</td>
<td>7</td>
<td>10.30</td>
<td>-6.17</td>
</tr>
<tr>
<td>58</td>
<td>7.05</td>
<td>4.13</td>
<td>2.90</td>
<td>7</td>
<td>10.30</td>
<td>-6.17</td>
</tr>
<tr>
<td>59</td>
<td>4.55</td>
<td>4.16</td>
<td>2.91</td>
<td>8</td>
<td>10.43</td>
<td>-6.27</td>
</tr>
<tr>
<td>60</td>
<td>4.55</td>
<td>4.16</td>
<td>2.91</td>
<td>8</td>
<td>10.43</td>
<td>-6.27</td>
</tr>
<tr>
<td>61</td>
<td>7.19</td>
<td>4.16</td>
<td>2.91</td>
<td>8</td>
<td>10.43</td>
<td>-6.27</td>
</tr>
<tr>
<td>62</td>
<td>7.19</td>
<td>4.16</td>
<td>2.91</td>
<td>8</td>
<td>10.43</td>
<td>-6.27</td>
</tr>
<tr>
<td>63</td>
<td>7.19</td>
<td>4.16</td>
<td>2.91</td>
<td>8</td>
<td>10.43</td>
<td>-6.27</td>
</tr>
<tr>
<td>64</td>
<td>7.19</td>
<td>4.16</td>
<td>2.91</td>
<td>8</td>
<td>10.43</td>
<td>-6.27</td>
</tr>
<tr>
<td>65</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>5.02</td>
<td>3.53</td>
<td>2.91</td>
<td>9</td>
<td>6.63</td>
<td>-3.10</td>
</tr>
<tr>
<td>69</td>
<td>5.02</td>
<td>3.53</td>
<td>2.91</td>
<td>9</td>
<td>6.63</td>
<td>-3.10</td>
</tr>
<tr>
<td>70</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>5.10</td>
<td>2.83</td>
<td>2.19</td>
<td>10</td>
<td>6.07</td>
<td>-3.24</td>
</tr>
<tr>
<td>74</td>
<td>2.77</td>
<td>2.26</td>
<td>2.02</td>
<td>10</td>
<td>3.45</td>
<td>-1.19</td>
</tr>
<tr>
<td>75</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>6.33</td>
<td>2.26</td>
<td>2.02</td>
<td>11</td>
<td>3.45</td>
<td>-1.19</td>
</tr>
<tr>
<td>77</td>
<td>6.33</td>
<td>2.26</td>
<td>2.02</td>
<td>11</td>
<td>3.45</td>
<td>-1.19</td>
</tr>
<tr>
<td>78</td>
<td>6.33</td>
<td>2.26</td>
<td>2.02</td>
<td>11</td>
<td>3.45</td>
<td>-1.19</td>
</tr>
<tr>
<td>79</td>
<td>2.39</td>
<td>1.53</td>
<td>1.34</td>
<td>12</td>
<td>2.47</td>
<td>-0.94</td>
</tr>
<tr>
<td>80</td>
<td>2.39</td>
<td>1.53</td>
<td>1.34</td>
<td>12</td>
<td>2.47</td>
<td>-0.94</td>
</tr>
<tr>
<td>81</td>
<td>2.39</td>
<td>1.53</td>
<td>1.34</td>
<td>12</td>
<td>2.47</td>
<td>-0.94</td>
</tr>
<tr>
<td>82</td>
<td>2.39</td>
<td>1.53</td>
<td>1.34</td>
<td>12</td>
<td>2.47</td>
<td>-0.94</td>
</tr>
<tr>
<td>83</td>
<td>2.39</td>
<td>1.53</td>
<td>1.34</td>
<td>12</td>
<td>2.47</td>
<td>-0.94</td>
</tr>
<tr>
<td>84</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>5.64</td>
<td>1.45</td>
<td>0.52</td>
<td>13</td>
<td>6.15</td>
<td>-4.69</td>
</tr>
<tr>
<td></td>
<td>2.39</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>----</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>87</td>
<td>2.39</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>88</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>2.37</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>95</td>
<td>2.37</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>96</td>
<td>2.37</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>97</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>15</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>2.37</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>99</td>
<td>2.37</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>100</td>
<td>2.37</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>101</td>
<td>2.37</td>
<td>1.40</td>
<td>1.34</td>
<td>14</td>
<td>1.70</td>
<td>-0.30</td>
</tr>
<tr>
<td>102</td>
<td>2.37</td>
<td>2.29</td>
<td>1.89</td>
<td>15</td>
<td>4.29</td>
<td>-2.00</td>
</tr>
<tr>
<td>103</td>
<td>2.37</td>
<td>2.29</td>
<td>1.89</td>
<td>15</td>
<td>4.29</td>
<td>-2.00</td>
</tr>
<tr>
<td>104</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>3.59</td>
<td>3.87</td>
<td>3.18</td>
<td>16</td>
<td>7.30</td>
<td>-3.43</td>
</tr>
<tr>
<td>107</td>
<td>3.59</td>
<td>3.87</td>
<td>3.18</td>
<td>16</td>
<td>7.30</td>
<td>-3.43</td>
</tr>
<tr>
<td>108</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>4.73</td>
<td>3.54</td>
<td>3.04</td>
<td>18</td>
<td>6.04</td>
<td>-2.50</td>
</tr>
<tr>
<td>111</td>
<td>4.73</td>
<td>3.54</td>
<td>3.04</td>
<td>18</td>
<td>6.04</td>
<td>-2.50</td>
</tr>
<tr>
<td>112</td>
<td>4.73</td>
<td>3.54</td>
<td>3.04</td>
<td>18</td>
<td>6.04</td>
<td>-2.50</td>
</tr>
<tr>
<td>113</td>
<td>4.73</td>
<td>3.54</td>
<td>3.04</td>
<td>18</td>
<td>6.04</td>
<td>-2.50</td>
</tr>
<tr>
<td>114</td>
<td>4.73</td>
<td>3.54</td>
<td>3.04</td>
<td>18</td>
<td>6.04</td>
<td>-2.50</td>
</tr>
<tr>
<td>115</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>116</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>117</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>118</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>119</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>120</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>121</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>122</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>123</td>
<td>2.70</td>
<td>1.41</td>
<td>1.54</td>
<td>18</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>124</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>&gt;6</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>5.42</td>
<td>3.47</td>
<td>2.63</td>
<td>19</td>
<td>7.70</td>
<td>-4.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>133</td>
<td>5.42</td>
<td>3.47</td>
<td>2.63</td>
<td>19</td>
<td>7.70</td>
<td>-4.23</td>
</tr>
<tr>
<td>134</td>
<td>2.74</td>
<td>1.59</td>
<td>2.64</td>
<td>19</td>
<td>-3.68</td>
<td>5.26</td>
</tr>
<tr>
<td>135</td>
<td>2.74</td>
<td>1.59</td>
<td>2.64</td>
<td>19</td>
<td>-3.68</td>
<td>5.26</td>
</tr>
<tr>
<td>136</td>
<td>3.49</td>
<td>3.06</td>
<td>2.67</td>
<td>20</td>
<td>5.03</td>
<td>-1.96</td>
</tr>
<tr>
<td>137</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>138</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>5.68</td>
<td>2.95</td>
<td>2.39</td>
<td>22</td>
<td>5.72</td>
<td>-2.78</td>
</tr>
<tr>
<td>141</td>
<td>5.68</td>
<td>2.95</td>
<td>2.39</td>
<td>22</td>
<td>5.72</td>
<td>-2.78</td>
</tr>
<tr>
<td>142</td>
<td>5.68</td>
<td>2.95</td>
<td>2.39</td>
<td>22</td>
<td>5.72</td>
<td>-2.78</td>
</tr>
<tr>
<td>143</td>
<td>5.68</td>
<td>2.95</td>
<td>2.39</td>
<td>22</td>
<td>5.72</td>
<td>-2.78</td>
</tr>
<tr>
<td>144</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>5.68</td>
<td>2.95</td>
<td>2.39</td>
<td>22</td>
<td>5.72</td>
<td>-2.78</td>
</tr>
<tr>
<td>146</td>
<td>X</td>
<td>&lt;3</td>
<td>&lt;3</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>5.68</td>
<td>2.93</td>
<td>2.36</td>
<td>23</td>
<td>5.79</td>
<td>-2.85</td>
</tr>
<tr>
<td>148</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>&lt;0.5</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>154</td>
<td>5.85</td>
<td>2.92</td>
<td>2.35</td>
<td>24</td>
<td>5.80</td>
<td>-2.88</td>
</tr>
<tr>
<td>155</td>
<td>5.85</td>
<td>2.92</td>
<td>2.35</td>
<td>24</td>
<td>5.80</td>
<td>-2.88</td>
</tr>
</tbody>
</table>