Title

Mapping Complex Protein Energy Landscapes with Covalent Labeling Methods

Author

Costello, Shawn Michael

Publication Date

2023

Peer reviewed|Thesis/dissertation

Mapping Complex Protein Energy Landscapes with Covalent Labeling Methods

By

Shawn M. Costello

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 Jamie H.D. Cate
Professor Andreas Martin
Doctor Corie Y. Ralston

Spring 2023

Abstract

Mapping Complex Protein Energy Landscapes with Covalent Labeling Methods

by

Shawn M. Costello

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Susan Marqusee, Chair

Proteins are large, complex macromolecules that play a wide variety of essential roles in living organisms. It has long been appreciated that the amino acid sequence of a protein encodes its three-dimensional structure, which is essential for biological function. It is becoming increasingly appreciated that protein structure is not static; proteins are dynamic molecules, occupying many conformations with varying populations on a broad range of timescales. This conformational ensemble can be thought of as an energy landscape, and be described using the language of kinetics and thermodynamics. To truly understand how proteins execute their broad set of functions we need to understand how these energy landscapes are encoded by protein sequence, how they determine protein function, and how they are influenced by biological environments.

Covalent labeling methods are ideal tools for answering these questions, as the chemical details of different covalent labeling reactions make them sensitive to protein structure, stability, and dynamics, and the temporal separation between labeling and detection facilitates the use of these methods on complex mixtures of proteins and other macromolecules. In this work, I use multiple covalent labeling methods to map the details of protein energy landscapes. First, I provide a background on protein conformational ensembles, their timescales, and on the covalent labeling methods used in this work. Second, I discuss my developments using a combination of hydroxyl radical footprinting mass spectrometry (HRF-MS) and chemical denaturation to extend our ability to measure protein global thermodynamic stability to a broader range of proteins and solution conditions. Third, I report on our use of hydrogen-deuterium exchange mass spectrometry (HDX-MS) to describe previously unknown conformational heterogeneity of the SARS-CoV-2 spike protein and detail how this heterogeneity is modulated by temperature, sequence, receptor binding, and interaction with antibodies. Finally, I describe work using a combination of labeling methods including HDX/MS, thiol labeling, and active site labeling to determine how the folding trajectory of the protein HaloTag is altered by translation. Lastly, I explain early efforts to extend these approaches to obtain more detailed structural information on the folding of proteins during translation.

Table of Contents

Acknowledgements

First, I want to thank my family — especially my mom Tracy, my dad Keith, and my sisters, Dianna and Jade. I know how lucky I am to have a family that is trusting and supportive of my career choices, even when those choices involved me moving far away from home. Thank you for your unwavering support and love.

I wouldn't have made it to graduate school without many generous mentors. I owe so much to all my teachers and peers at Miami Dade College, especially my first scientific mentor, James Ley. Thank you for teaching me that science was for everyone and being the first person to tell me I could succeed. I'm sorry I didn't do more NMR after leaving MDC; I promise I'll try and do more in the future. I also would not be here without all my professors, mentors, and classmates in the Biophysics department at Johns Hopkins University. I came to Hopkins knowing nothing about proteins and left wanting to spend the rest of my career studying them. I owe most of that to Pat Fleming, Ashlee Plummer, and Karen Fleming. Pat, I was terrified when you, very seriously, asked to "talk to me after class," but that conversation led to me joining Karen's lab and ultimately going to graduate school. Thank you for being such a dedicated and engaging teacher in the classroom and the lab. Ashlee, I learned so much from you. Now that I'm on the other side of graduate school, I can see more clearly how you went above and beyond as a mentor. Thank you for being so patient and selfless. Karen, I learned more about how to be a scientist from you than anyone else. I have so much respect for your excitement, your values, and your dedication. Thank you for being such a caring mentor.

Of course, I want to thank my advisor, Susan Marqusee. Thank you for always letting me explore my interests. I know it was a meandering path, but I'm proud of what I accomplished, and I owe so much of it to you. Thank you for all the time and energy you put into the unsung work behind the scenes that makes science possible. In your lab, I didn't just learn how to do research; I learned how to be a good scientific citizen. I hope to live up to your example in my future scientific career. Most of all, thank you for being as nerdy about proteins as I am.

Thank you to everyone at Berkeley who has helped me over the past seven years. First, I want to thank my thesis committee members, Jamie Cate, Corie Ralston, and Andy Martin, for being so understanding of my hectic and shifting timetable. I've been so fortunate to work with and learn from so many incredible scientific collaborators at Berkeley and around the country. I want to especially thank Sayan Gupta, who taught me everything I know about X-ray footprinting. Thank you for always accommodating my schedule and the many last-minute changes. Lastly, I want to thank Kate Chase for being the glue that holds the biophysics program together. As of the

submission of this thesis, you have sent me 4151 emails. I can't promise I have read every one of them, but I hope there will be more.

I want to thank all the current and former members of the Marqusee lab that I've worked alongside over the years. Thank you to the lab members who were here when I started, Shion Lim, Eric Bolin, Emily Guinn, Ha Truong, Madeleine Jensen, Emma Carroll, and Helen Hobbs. I learned so much from your example. I want to especially thank Avi Samelson for being such a great first rotation mentor and for continuing to be a mentor and friend. I also want to thank Brendan REDACTED for many things. Thank you, of course, for all the stocks, gels, ordering, and protein preps, but also, thank you for showing so much solidarity. I won't ever forget it. I also want to thank all the lab members who joined after me, Sara Volz, Miriam Hood, Tram Nguyen, Darren Kahan, Eva Gerber, Johanna Lindner, and Carol Mendonça. Thank you for making every day in lab brighter. I want to especially thank Naomi Latorraca, Sophie Shoemaker, and Natalie Dall. Naomi, whenever I have a problem, you are the first person I go to for advice. I have so much respect for you as a scientist and person. Sophie, you are the best co-author anyone could ask for. Our paper wouldn't have happened without your hard work and brilliance. Natalie, thanks for being the person I could always talk (complain) to. You are such a caring, honest, and fun person, and I'm lucky to have you as a friend. Every one of the (many) rough days in lab was made easier, and every great day was more joyful, because of all of you.

Lastly, I want to acknowledge how lucky I am to have made many incredible friends. Thank you to all the friends mentioned above, along with Mina Farag, Quenton Bubb, Jill Hakim, Alex Crits-Christoph, Rachel Viqueira, Jonathan Liu, Dagan Marx, Erika López-Alfonzo, and Josiah Johnson. Jill and Alex, besides being two amazing friends, you are the best roommates and two of the most brilliant scientists I know. Lastly, I want to thank Charlotte Nixon. You fall into so many categories — mentor, coworker, roommate, and friend. You helped me so many times during the most challenging moments of graduate school. I will treasure your friendship forever. To all my friends, I am constantly amazed by your brilliance, kindness, and humility. You have all made me a better person. Thank you.

Chapter 1

Introduction

The bulk of this dissertation consists of biophysical studies leveraging covalent labeling methods to understand protein folding, stability, and dynamics. The following introduction serves to provide a brief background of the protein folding field, the motivation for this work, and to help those not in the field better appreciate the details presented in the following chapters.

1.1 Explaining protein function with conformational ensembles

Proteins are life's primary tool for biological function — with the capability of catalyzing chemical transformations, providing cellular structure, and communicating and receiving signals, amongst many other tasks. Proteins accomplish this wide variety of functions despite all being linear polymers of the same set of 20 of amino acids. Each protein is defined by its specific sequence of amino acids. Understanding how this linear sequence of amino acids can encode such diverse functions has been a central goal of biology for the last half-century. Recently, great strides have been made in our ability to predict a three-dimensional structure for a given protein sequence.[1,2] While protein structure is highly informative, knowing this single structure for a protein is not sufficient to understand its mechanism of function. Proteins are not static structures, but highly dynamic molecules, sampling many conformations, with varying stability, on various timescales (Fig. 1-1). This conformational ensemble ultimately dictates protein function — fluctuations in enzymes are essential for substrate binding, catalysis, and product release; allosteric communication relies upon selective stabilization of conformations within the ensemble as the result of binding of a ligand; transitions to unfolded and partially folded states dictate the turnover rate of proteins by cellular quality control machinery, etc.[3,4] Perturbations to proteins, such as the binding of a metabolite, change in environmental conditions, or the introduction of disease-causing missense mutations, do not just affect a protein's structure; they affect the relative energetics of these conformations and the rates of interconversion between them. In other words, perturbations alter protein energy landscapes in complex ways. If we truly want to predict protein function from the sequence alone, we must understand protein energy landscapes at this mechanistic level.

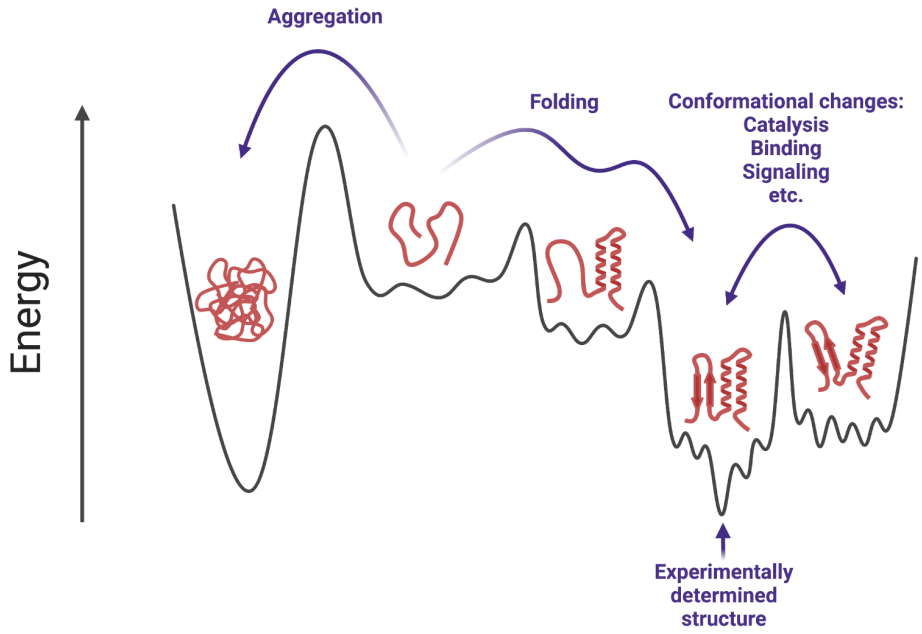


Figure 1-1. Protein energy landscape. A reaction coordinate (x-axis) depicting the energy (z-axis) of various conformational states. The population of each state is determined by the relative energy of its well, and the rate of transitioning between states is determined by the heights of barriers between wells. Folding and aggregation flux are controlled by kinetic barriers. Biological function frequently relies on transitions between conformations. Functional conformational states typically have conformational substates, represented by the smaller wells within each larger well. Structural tools typically report only on one conformation.

1.1.1 Complex solution conditions (e.g., Translation)

Protein energy landscapes are sensitive to the physicochemical environment; conformational ensembles are sensitive to, for example, temperature, pH, and pressure.[5–7] In many cases, this response is leveraged by cells, allowing organisms to sense and respond to these stimuli.[8–10] Many other qualities of biological environments perturb protein energy landscapes. The cellular interior is highly crowded with other macromolecules in addition to small molecules, ions, and metabolites, many of which are known to interact with proteins and affect their conformational ensembles.[11,12] Also, biological environments are not at equilibrium, the abundance of these molecules can change over time, and proteins are constantly being created by translating ribosomes and destroyed by the cellular quality control machinery.[13] This non-equilibrium nature means that many proteins do not exhaustively sample their entire energy landscapes but can be kinetically trapped in conformations or prevented from accessing certain states by cellular folding machinery.[14–16]

One fascinating difference between protein folding in the cell compared to proteins studied in isolation in a test tube is the potential for folding during translation by the ribosome. All proteins are synthesized by ribosomes, which decode messenger RNA to synthesize nascent proteins one amino acid at a time. This process is relatively slow, occurring at around 1-20 amino acids per second, meaning the complete translation of a typical protein can take up to several minutes.[17]

Proteins begin to fold and interact with the cellular machinery during this process.[18] The varied physicochemical environments experienced by proteins and the sensitivity of protein energy landscapes to the physicochemical environment emphasize the importance of studying proteins in as native an environment as possible.

1.1.2 Limitations of existing approaches

The limitations of existing methods have constrained our understanding of protein energy landscapes. Ideally, we would be able to probe the energy landscape for any given protein in its native cellular environment. Traditional methods, like circular dichroism (CD) or tryptophan fluorescence, require large quantities of protein at micromolar concentrations, challenging for many, if not most, proteins. Additionally, traditional methods often require high purity. In the case of spectroscopic methods, all biomolecules in a sample produce a signal, making it impossible to specifically monitor a protein of interest within a more complex mixture. Lastly, for rigorous thermodynamic characterization, proteins must fold and unfold reversibly and ideally are two-state (only occupy a folded and unfolded confirmation, without any populated partially folded conformations). As a result, most understanding of protein energy landscapes comes from studies of a small subset of proteins that tend to be smaller than the average protein and are much more likely to be monomeric, cytosolic, and water-soluble.[19] The complexity of biological systems and the limitations of existing approaches demands the development of new methodologies if we are to understand protein energy landscapes in their native environments.

1.2 Covalent labeling approaches

Covalent labeling methods have several advantages over traditional spectroscopic methods. A key advantage is a temporal separation of "encoding" and "reading" structural information. For example, in the case of tryptophan fluorescence, the information is "encoded" by the absorbance of light and "read out" as the emission of a different wavelength of light. In this example, and in most spectroscopic experiments in general, the information is encoded and read essentially instantaneously. For covalent labeling methods (specific methods discussed in further detail below), the information is "encoded" by a covalent modification, which is stable for minutes to hours or even longer after the experiment, allowing the experimenter to "read out" the information at a later time.

The long timescale between the "encoding" and "reading" provides multiple benefits. The primary benefit is that the protein of interest can be separated from complex mixtures of biomolecules, greatly relaxing the purity constraints from traditional spectroscopic methods. Additionally, the "reading" of the encoded information can be done by many methods that detect molecules with high sensitivity but are not amenable to native solution conditions (i.e. aqueous, neutral pH, moderate ionic strength, etc.). This increased sensitivity greatly relaxes the quantity of sample needed to perform a single measurement, and removing the requirement for native solution conditions expands the repertoire of physical methods that can be applied. The most frequently used method for covalent labeling studies of proteins is reversed-phase liquid chromatography coupled to mass spectrometry (LC/MS). Reversed-phase LC/MS involves

separating mixtures of biomolecules under acidic conditions on a hydrophobic column with increasing amounts of organic solvent, followed by ionization/desolvation into the gas phase and detection by a variety of mass detectors. Frequently, in LC/MS, proteins are first digested into fragments (i.e., peptides), providing local rather than global information. This process of digesting proteins into peptides and analyzing these fragments by mass spectrometry is called "bottom-up" mass spectrometry.[20] Local information can also be obtained by performing what is called "top-down" mass spectrometry on intact proteins, fragmenting them within the mass spectrometer, and analyzing the mass of the resulting fragments.[21,22] Which approach is best depends on the particular experimental question, whether the sample is amenable to one or the other approach, and what instrumentation is available. The ability to obtain local structural information is essential for studying larger non-two-state proteins with more complex energy landscapes. Many covalent labeling methods exist, each with its own accessible questions and logistical idiosyncrasies. In this thesis, I will discuss three: hydrogen-deuterium exchange, hydroxyl-radical footprinting, and generalizable protein-ligand binding. However, before discussing what is unique to each of these methods, it is helpful to begin by discussing the kinetic theories and assumptions that are shared between them.

1.2.1 Theory of chemical labeling

The theoretical framework for interpreting covalent labeling experiments is generalizable, despite the chemistry for different covalent labeling reactions varying dramatically. This framework, first developed by Linderstrøm-Lang and colleagues for the interpretation of backbone hydrogen exchange within proteins, categorizes a reactive site in a particular conformation of a protein as either "open" or "closed" where the open state is accessible and capable of reacting with the chemical modifying agent, and the closed state is not (Fig. 1-2).[23–28] Like all models, this framework is a clear simplification, yet it is sufficient for describing most cases. The fundamental assumption that a site within a protein can be described by an "open" accessible state and a "closed" inaccessible state is biophysically consistent with our understanding of globular proteins. Globular proteins are folded into a well-defined structure, with hydrophobic residues in well-packed cores that are inaccessible to solvent, and with a large fraction of backbone amide groups engaged in hydrogen bonds forming secondary structures. While buried or engaged in hydrogen bonds, these residues or backbone functional groups are sterically unable to react with covalent modifying agents. However, as proteins are dynamic, occupying many conformations with varying populations, even buried residues or hydrogen bonds will transiently open and become exposed to solvent, allowing for labeling.[29–31]

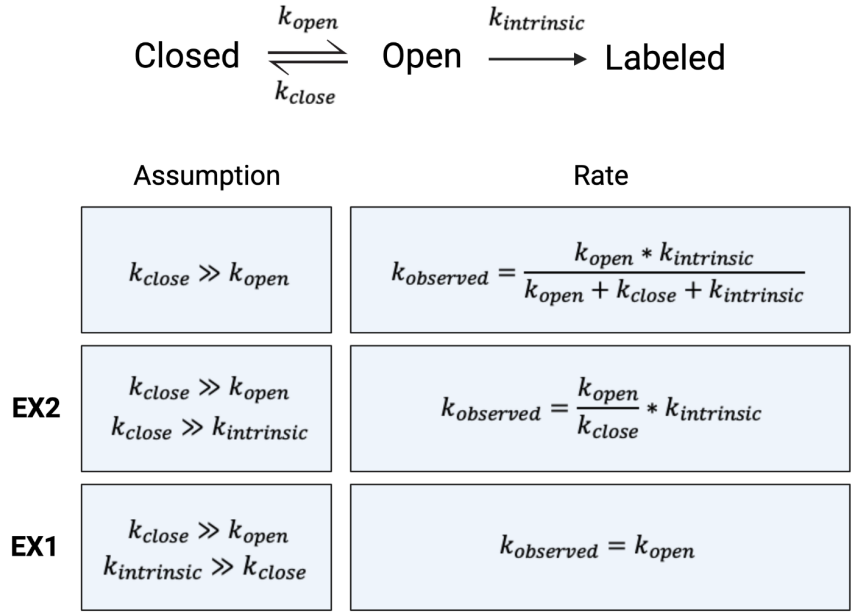


Figure 1-2. Kinetic scheme for covalent labeling. Assuming there is an open and closed state in equilibrium, where only the open state is capable of reacting with the label, one can derive expressions relating the observed rate of labeling to the rate constants in this kinetic scheme. Under different kinetic regimes described on the left, expressions simplify to report on either the energetics or kinetics of the opening reaction.

Using this simple framework, one can derive expressions relating the observed rate ($k_{observed}$) of modification to the rates of conformational interconversion (k_{open} and k_{close}), and the "intrinsic" rates describing the chemistry or reactivity of completely accessible modification sites ($k_{intrinsic}$, also frequently referred to as $k_{chemical}$) (Fig. 1-2). Importantly, the terms in this expression, k_{open}, and k_{close}, are quantitative, physical descriptions of the protein conformational ensemble. Each rate constant describes the dynamics of a conformational change, and the ratio (k_{open}/k_{close}) describes the equilibrium populations (K_{open}) which is directly related to the free energy of that conformational change ($\Delta G = -RTln(K_{open})$). This derived expression assumes that k_{open} is much smaller than k_{close} (Fig. 1-2). This is equivalent to saying that the closed state must be thermodynamically favored (more populated at equilibrium) than the open state. Again, since folded proteins are generally stable with a large energetic gap between any other accessible conformations, this assumption is generally valid (with many exceptions that are not discussed in this work).

The expression defined in Figure 1-2 is complex, and the exact determination of the parameters of interest (k_{close}, k_{open}, K_{open}) from $k_{observed}$ is not trivial. However, there are several kinetic regimes that simplify this expression. If the $k_{intrinsic}$ is much slower than k_{close}, then the observed rate will depend on the equilibrium populations of the so-called closed and open states and the intrinsic rate ($k_{observed} = K_{open} * k_{intrinsic}$) (Fig. 1-2). This behavior is frequently referred to as EX2 behavior.[28,32,33] If, instead, $k_{intrinsic}$ is very fast, much faster than k_{close}, then the observed rate will simply be the rate of transitioning to the open state ($k_{observed} = k_{close}$) (Fig. 1-2). This behavior is frequently referred to as EX1 behavior.[28,32–35] Both regimes are informative; in EX2, when the timescales of labeling are much slower than conformational changes, the modification rates

report on the equilibrium properties of the conformational ensemble, and in EX1, when labeling is much faster than conformational changes, the modification rates depend on the kinetic or structural properties of the ensemble. Whether a protein falls into one regime or another depends on the specific protein, the solution conditions, and on the chemistry of the labeling method.

1.2.2 Continuous vs. pulsed labeling

There are two general categories of labeling experiments, referred to as "continuous labeling" and "pulsed labeling." Continuous labeling experiments involve initiating the labeling reaction and monitoring the degree or fraction of molecules labeled over time. From these experiments, under the kinetic regimes described above, one can learn about the dynamics (k_{open}) or energetics (K_{open}) of the system (Fig. 1-2). Often, the experiment is repeated under different conditions, typically the addition of a ligand or introduction of a mutation, to see how these physical parameters change, providing insight into the mechanism of action, and structural changes, for a given perturbation.[33,36–38]

Pulsed-labeling experiments involve the selection of a short timescale pulse for the covalent labeling reaction, and then carrying out this pulse as a function of some other variable, often the time course of a reaction.[33,39] While the kinetics of labeling explained above still apply, typically the goal of these experiments is not to determine any parameters relating to the labeling reaction itself, but to provide structural and mechanistic insight into the reaction being probed. For example, one could monitor the formation of structural elements during protein folding by first initiating refolding (typically by dilution from chemical denaturant) and then applying a short pulse of labeling after different refolding durations.[40–48] This experiment requires that the pulse length be much shorter than the rate of the reaction of interest. In the case of the refolding of RNase H, which occurs on the second timescale, a hydrogen exchange pulse length of 10 milliseconds was used, whereas, for the conformational change in influenza hemagglutinin, which occurs in minutes, a hydrogen exchange pulse length of 10 seconds was used.[40,49]

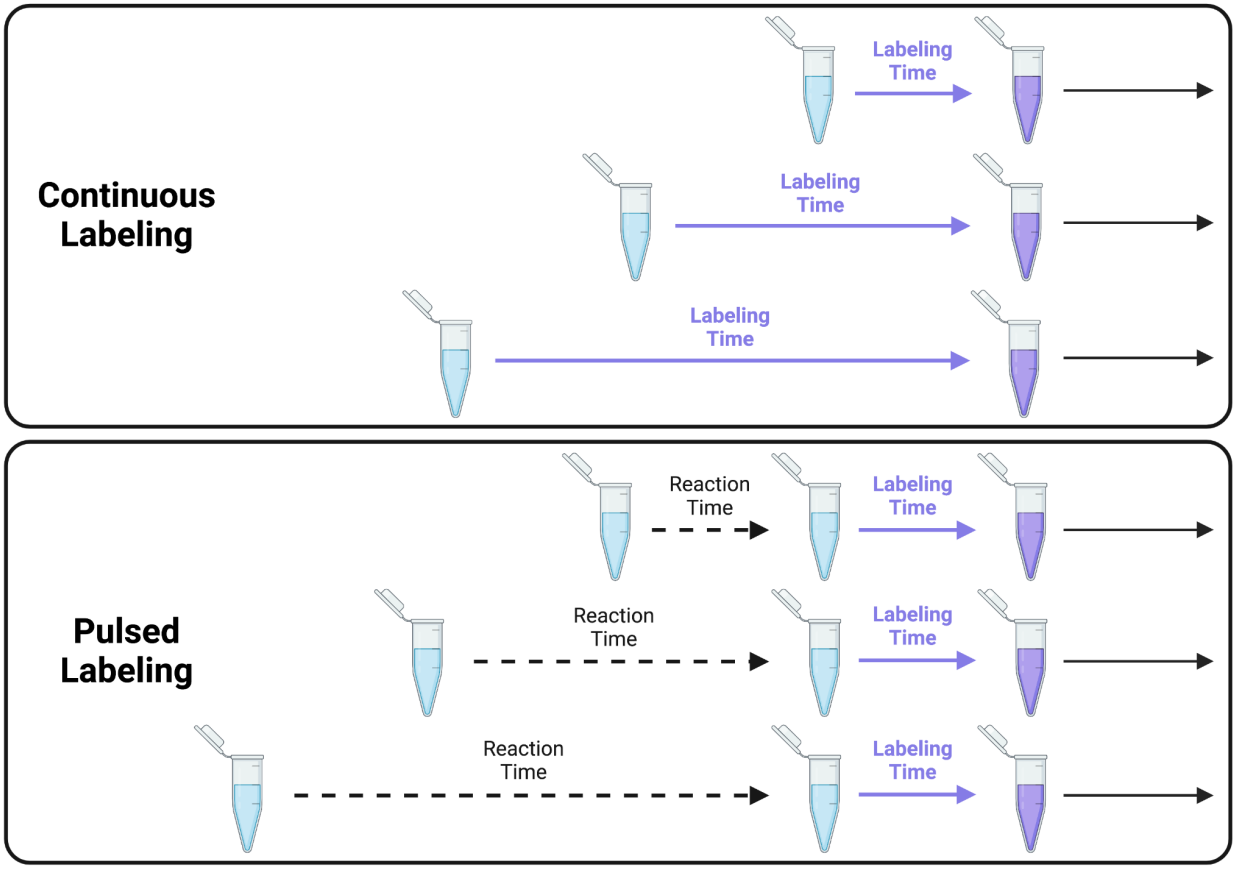


Figure 1-3. Two main categories of kinetic labeling experiments. In continuous labeling, the labeling reaction is initiated and the degree of labeling is monitored over time, typically by quenching after various times and measuring the degree of labeling at each time point. In pulsed labeling, a reaction is initiated, and at various times during that reaction, a short pulse of labeling is performed, followed by quenching and measurement of the degree of labeling.

1.2.3 Hydrogen deuterium exchange

Hydrogen Deuterium Exchange (HDX) is a method that reports on the solvent accessibility and stability of backbone hydrogen bonds in proteins (Fig. 1-3). Protein secondary structures, such as alpha helices and beta-sheets, are defined by specific patterns of backbone-backbone hydrogen bonds. These non-covalent bonds are between the carbonyl oxygen of one peptide bond and the amide hydrogen of another. Although the amide hydrogen is covalently attached to the amide nitrogen, it is labile, meaning it can dissociate as a free proton and be replaced with another proton from the aqueous solvent. This exchange reaction can be catalyzed, with varying efficiencies, by acid (H+), base (OH-), or water (H_2O), making the exchange reaction sensitive to pH. The intrinsic rate ($k_{intrinsic}$) of the exchange reaction depends on the pH and the amino acid side chains from the adjacent residues. These intrinsic exchange rates have been thoroughly characterized as a function of pH, temperature, and ionic strength.[50–53] As described above, in the context of a structured protein, this exchange reaction can be slowed dramatically due to the inability for exchange when the amide hydrogen is engaged in a hydrogen bond or when that hydrogen is not exposed to the solvent (the so-called ‘closed’ conformation, see above). As a result, the observed exchange rate depends on sampling conformations where that specific amide

hydrogen is exposed to solvent and not engaged in a protein-protein hydrogen bond (see Linderstrøm-Lang formalism described in section 1.2.1).

The two primary methods for monitoring deuterium uptake during an HDX experiment are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). In HDX-NMR, usually a heteronuclear single quantum coherence spectroscopy (HSQC) experiment is used since each amide nitrogen-hydrogen pair produces a unique peak in a two-dimensional spectrum.[54,55] Since this NMR experiment detects hydrogen and not deuterium, the HDX reaction results in a loss of intensity of this peak. Therefore, the observed exchange rate is calculated by monitoring the peak intensity (or volume) for each amide over exchange time. The primary advantage of this method is that, under ideal conditions, each amide produces a unique signal so that each site can be monitored independently. However, this also requires knowledge of the assignment of each peak in the HSQC spectrum. Large proteins are more challenging due to increased spectral crowding and slower tumbling in solution. Practically, this limits HDX-NMR to proteins smaller than 35 kDa. NMR detection also has a limited range of viable solution conditions and protein concentrations. For these reasons, MS has become a primary method for monitoring HDX.

In an HDX-MS experiment, after the exchange reaction, samples at each exchange time are quenched by reducing the pH and temperature. Typically, proteins are then digested into fragments (peptides) under quenched conditions with an acid-active protease, separated by reversed-phase liquid chromatography (RPLC), ionized, and injected into a mass spectrometer where the mass-to-charge ratio for each peptide is measured. Since deuterium is one Dalton heavier than hydrogen, the average mass of the peptide is directly related to the number of hydrogens exchanged for deuterium. The primary advantage of this method is that it is more amenable to larger proteins or mixtures of proteins. However, this is at the expense of site-specific resolution. MS monitors the mass increase of a peptide containing multiple amides, making the resulting deuteration time course a sum of the individual exchange rates within that peptide. A high degree of overlap between the observed peptides sometimes allows for calculating individual rates, but achieving the necessary degree of overlap in practice is rare.[40,56–58] Recent advances in top-down HDX-MS allow for site resolved measures of deuteration, although there are many limitations as to which proteins are amenable. Despite this, the information provided by individual peptides is still local and, therefore, sufficient for many experimental questions.[22,59] Also, compared to HDX-NMR, HDX-MS requires less total sample and is easier to perform at lower concentrations, further increasing the types of protein systems amenable to HDX.

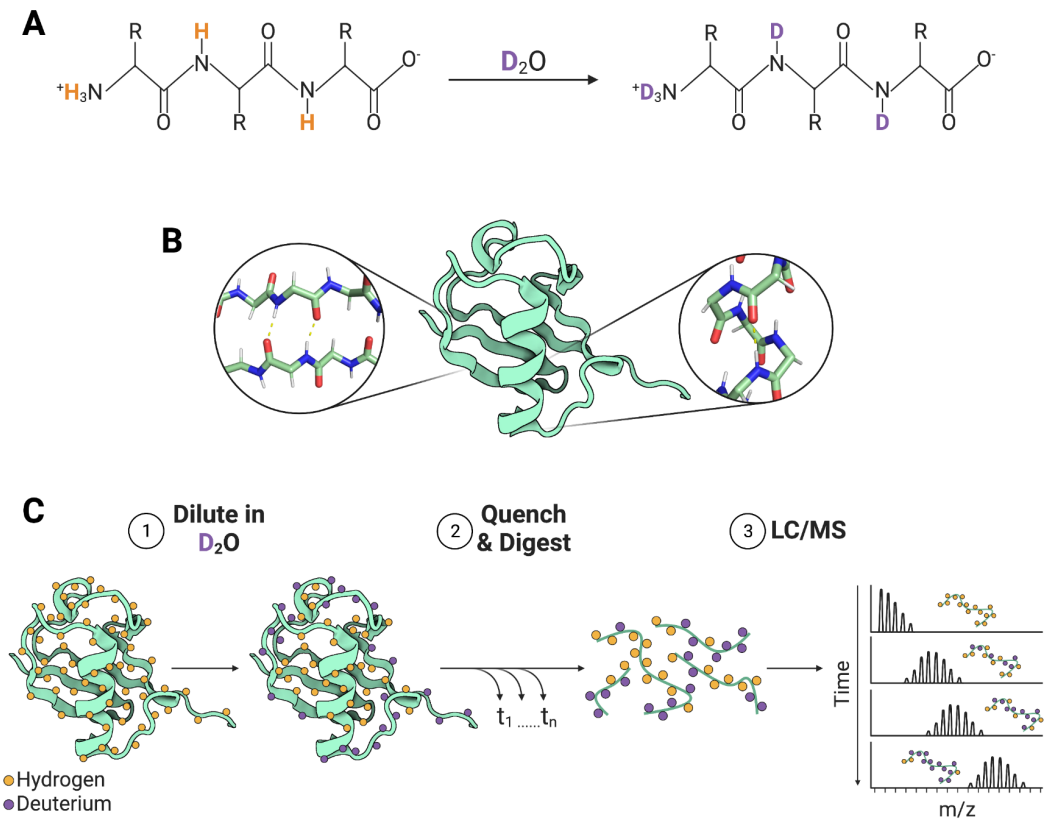


Figure 1-4. Mechanism and workflow for HDX. (A) Backbone hydrogen exchange reaction show for a generic protein backbone (side chains represented as R groups. Amide hydrogens are labile and when exposed to deuterated water (D_2O) will be replaced with deuterium. (B) Protein ribbon diagram with insets showing backbone amide hydrogen bonding in beta sheets (left) and alpha helices (right). (C) Workflow for continuous labeling HDX-MS. Protonated proteins are diluted in D_2O (1). After varying labeling times samples are quenched and proteins digested (2). Peptide mass distributions are then measured by LC/MS (3).

1.2.4 Hydroxyl Radical Footprinting

Hydroxyl radical footprinting (HRF) is a covalent labeling method that reports on the solvent accessibility of amino acid side chains within a protein (Fig. 1-4).[60] Hydroxyl radicals are highly reactive molecules and, when mixed with protein samples, will rapidly react with amino acid side chains.[61] The reaction chemistry with different side chains is highly variable; the most common outcome is the net addition of a single oxygen atom, resulting in a mass increase of 16 daltons, though other mass additions or losses are observed to a lesser extent.[61] The relative reactivities of the 20 canonical amino acids and the types of the resulting modifications have been characterized in detail.[60] The critical aspect of this experiment that makes it useful as a structural tool is that the hydroxyl radicals are of similar size to water molecules and are water-soluble. As a result, the observed modification rates of individual side chains relate to their solvent accessibility.[62] HRF has been effectively used to study protein-ligand binding, protein-protein interactions, and protein conformational changes.[63–65]

Two of the most common methods for generating hydroxyl radicals are radiolysis of water and laser photolysis of hydrogen peroxide. Radiolysis of water can be performed on the millisecond timescale using high-flux X-rays from a synchrotron light source.[66] This method requires no chemical additives, as the radicals are generated directly from the water molecules in the solution.[61] Recent advances in synchrotron beamline engineering and sample exposure apparatus design have reduced necessary labeling times to the microsecond range.[65,67,68] Laser photolysis of hydrogen peroxide, commonly referred to as FPOP (for fast photochemical oxidation of proteins), requires the addition of millimolar concentrations of hydrogen peroxide followed by exposure of the sample to a 248 nm laser.[69] With the addition of radical scavengers, FPOP can reach microsecond labeling times.[70] Recent advances in instrumentation have allowed for in-cell FPOP (IC-FPOP).[71–75] Additionally, recent commercial instruments using plasma photoylsis of hydrogen peroxide have lowered the technical barriers for HRF and will likely encourage the more widespread use of the method.[76,77] The selection of synchrotron-based x-ray radiolysis vs. laser photolysis depends on whether the system of interest can tolerate the buffer requirements imposed by each method and by the experimenters' access to instrumentation.

Observed modification rates in an HRF experiment can be compared to intrinsic modification rates to generate a protection factor ($k_{intrinsic}/k_{observed}$), as commonly used in HDX analysis.[78] Unlike HDX, this analysis does not result in parameters directly related to the energy landscape. In HDX, under EX2 conditions, the protection factors are related to the free energy difference (ΔG) between the closed and open conformation. In HRF, analogous EX2 conditions are not achievable for several reasons. First, hydroxyl radicals are highly reactive, resulting in fast intrinsic labeling rates. Second, hydroxyl modifications are perturbative, changing the structure of the protein; short labeling times prevent proteins from changing conformation during the experiment. Both factors result in HRF being most effectively used as a pulsed-labeling experiment. HRF protection factors are also challenging to measure; achieving site-specific resolution by HRF-MS requires high detection sensitivity and chromatographic resolution. Despite this, when measurable, HRF protection factors correlate with structural parameters and are, therefore, a useful structural probe.[78] When site-specific resolution is not achievable, similar to most HDX-MS experiments, comparing peptides across different conditions is sufficient for many experimental questions. In summary, HRF-MS provides a structural snapshot where protein side-chain solvent accessibility changes are highly correlated with changes in the observed modification rates.

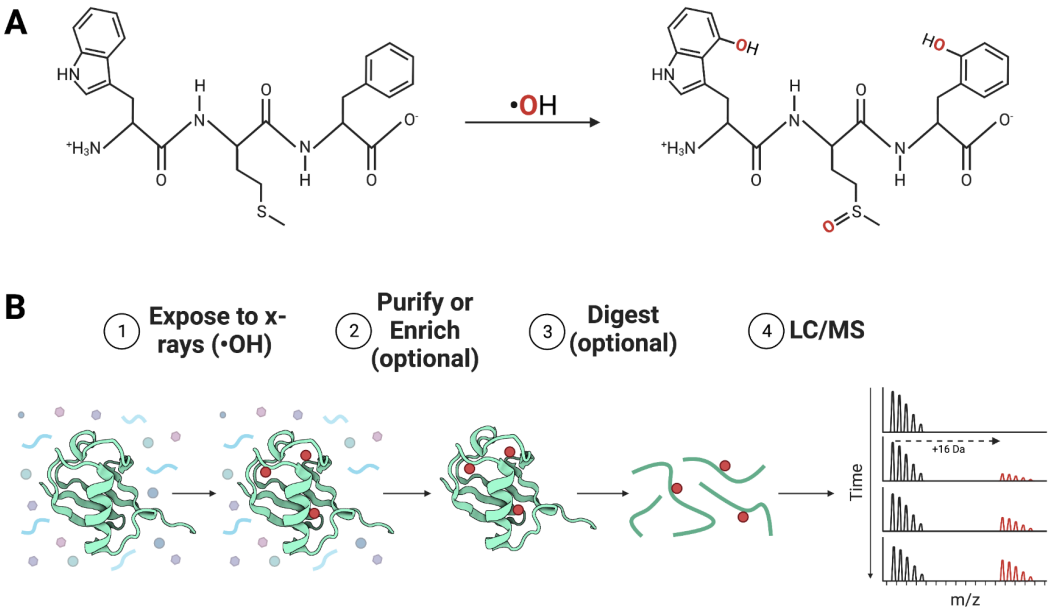


Figure 1-5. Mechanism and workflow for HRF. (A) Protein side chains are oxidized when exposed to hydroxyl radicals. (B) Workflow for HRF-MS. Protein samples are exposed to x-rays, which generate hydroxyl radicals from water (1). Proteins may be purified post-labeling (2) and digested (3). The degree of oxidation is then measured by detecting and quantifying both the unoxidized and oxidized proteins or peptides with LC/MS (4).

1.2.5 Ligand binding

As explained above, the formalism used for covalent labeling is highly generalizable. While HDX and HRF differ in what they report on (hydrogen bond stability vs. side chain solvent accessibility), the underlying logic of a modification rate relating to the underlying conformational distribution is the same. This same formalism can be extended to many other diverse reactions that modify proteins, such as cysteine modification or active-site labeling. One limitation of HRF and HDX is that both methods are non-specific; hydrogen exchange and hydroxyl radical reactions will occur on all biomolecules in the solution and at many sites within each biomolecule. While HRF and HDX have been employed in complex solutions, including in cells, these factors make data analysis quite challenging. A specific and/or bio-orthogonal reaction, where the reactivity is specific to the protein of interest and not reactive with any other biomolecules in solution, makes monitoring the protein of interest much more straightforward.[79] One way to generate a bio-orthogonal reaction is to rely on the activity of a specific enzyme or the highly specific and tight binding of a ligand. Several proteins have been engineered that covalently bind to a specific, non-biological ligand; two examples are SnapTag and HaloTag.[80–82] The activity of SnapTag has been used to monitor protein translation, and the activity of HaloTag has been used to monitor protein folding, as described later in this thesis.[83,84] While covalent attachment of the ligand is helpful and opens up certain experimental options, it is not necessary. Non-covalent binding can also be leveraged if the binding kinetics and affinity fall into the time regime of interest.

1.3 Timescales

As Howard Berg said, life at the cellular scale is "in constant riotous motion."[85] This is as true for molecules as it is for cells, and the timescales of these motions dictate which methods can be used. In studying protein energy landscapes, there are three different timescales to consider (i) the timescales of interconversion between conformations, (ii) the timescales accessible by the experimental method (for any method, not only covalent labeling methods), and (iii) the timescales of biological processes like synthesis, degradation, cell division, etc. Understanding how (i) and (ii) relate to each other is critical in designing covalent labeling experiments, as evidenced by the above discussion of kinetic regimes. Below is a comparison of the timescales of various protein conformational changes, how those relate to biological timescales, and what timescales are accessible with the covalent labeling methods discussed here.

1.3.1 Timescales of protein conformational transitions

Proteins are not static structures; they have many internal degrees of freedom and, like all molecules, are subject to thermal fluctuations. The time and length scales of protein internal motion both span orders of magnitude. This range extends from the Ångstrom scale vibrations of individual bonds on the scale of femtoseconds (10^{-15} seconds) to tens of nanometers global, concerted domain motions that can be as slow as tens of seconds (10^{1} seconds).[4] In the middle of these 16 orders of magnitudes is an incredible richness of molecular processes including, for example, the dissociation of individual non-covalent bonds, the isomerization of proline, and the rotameric transitions of side chains. Fraunfelder, Sligar and Wolynes coarse-grained this broad distribution of length and timescales into three "tiers".[3] The smallest scale tier consists of picosecond (10^{-12} seconds) local, atomic scale motions, such as the rotation of a methyl group on the surface of a protein. The largest scale tier consists of microsecond (10^{-6} seconds) to millisecond (10^{-3} seconds) collective, global protein conformational changes, such as the closing of Adenylate Kinase domains around a substrate. Of course, these tiers are nested, meaning a protein may have multiple conformational states defined by domain movements that interconvert on the millisecond time scale, and each of those conformational states will have many conformational substates defined by side chain rotameric transitions (Fig. 1-1). Because of the range of timescales, and the varying physical nature of these processes, the definition of a conformational transition is highly contextual and makes the selection of an appropriate physical method highly important (see below).

One important category of protein conformational change is folding, or the process by which a protein attains its native structure beginning from an unstructured "unfolded" state. While Anfinsen famously demonstrated that proteins can spontaneously reach their native state, Levinthal pointed out that the ability for proteins to spontaneously fold into native structures is surprising, as the number of degrees of freedom for a protein chain is so large.[86–88] While much is left to learn about the physics of protein folding, it is generally accepted that proteins overcome this apparent statistical impossibility by forming local structures that "guide" the protein toward the native state.[89–91] The timescales for these local structural transitions are on the oder of 10^{-8} seconds for hydrophobic collapse, 10^{-8}-10^{-6} seconds for helix-coil transitions, and 10^{-6} for

beta-hairpin closure.[92–96] The timescales for the acquisition of native structure from an unfolded state ranges even more dramatically. From a set of kinetically two-state reversible proteins, the timescales of folding (represented here by the average time to fold or unfold, τ_F and τ_U) ranges from 10 microseconds to 10 seconds for folding, and from 10 milliseconds to days for unfolding (Fig. 1-5).[97] This set of proteins is biased towards small well-behaved model systems, so the range of folding and unfolding timescales for all proteins is likely much larger.

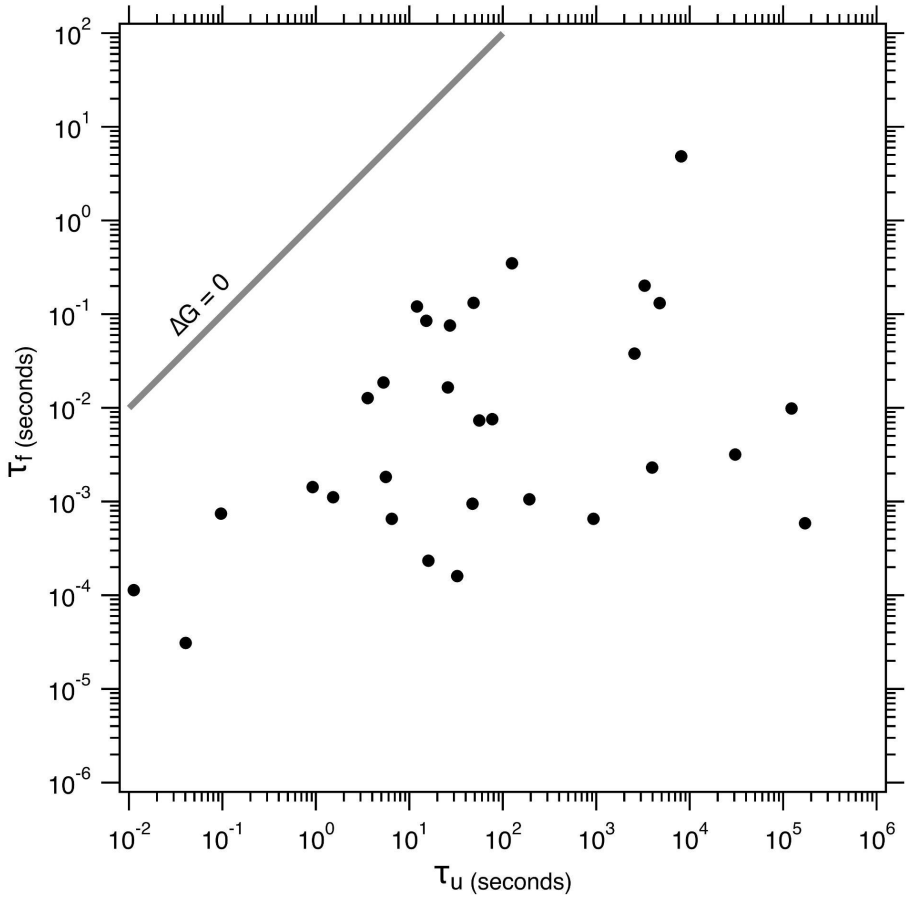


Figure 1-6. Distribution of folding and unfolding rates. A plot of the average folding time (τ_F) and unfolding time (τ_U) for a set of experimentally characterized two-state globular proteins.[97] Each point represents a single protein.

1.3.2 Timescales of biological processes

Translating the timescales of protein conformational transitions to biological function requires that the rates be placed in the context of various biological timescales. As discussed above, conformational changes, ranging in scale from side-chain rotamer transitions to large domain motions, occur on timescales much faster than seconds. However, many biological processes are much slower. Cell division for a rapidly dividing bacterium, molecular motors traversing cells, and transcription of single genes all take minutes.[98] Two important cellular processes for proteins are synthesis by the ribosome and degradation by cellular machinery. As discussed above, translation in bacteria occurs at a rate of 1-20 amino acids per second.[17] For an average sized protein (~150 amino acids), translation would take between 7.5 seconds and 2.5 minutes.

Translocation of a protein by the ClpXP bacterial degradation machinery occurs at nearly 80 amino acids per second, meaning translocation of an average sized protein would take a couple of seconds.[99] Because these processes are orders of magnitude slower than the picosecond to millisecond dynamics of proteins, one can conclude that proteins are thoroughly exploring their energy landscape during these processes. As a result, an increasing number of studies are being performed to understand how protein energy landscapes are affected by the biological machineries they interact with.[100–105]

Global folding and unfolding timescales, which are typically slower than other conformational transitions, must also be placed in the context of biological protein lifetimes. For instance, take a yeast protein that folds on the millisecond timescale and unfolds on the minute time scale. It will stably populate the folded state (K_{eq} ~= 0.001s/100s = 10^{-5}, ΔG = -7 kcal/mol) but will likely unfold and refold many times during its lifetime in the cell, as the average protein cellular lifetime (average time from synthesis to degradation) in yeast is about 50 minutes.[106] Take a second example, another yeast protein that folds on the second timescale and unfolds on the order of days. It will also stably populate the folded state, with the same stability as the previous example (K_{eq} ~= 1s/100000s = 10^{-5}, ΔG = -7 kcal/mol); however, once folded, it will be unlikely to unfold during its cellular lifetime. This has important implications for which biological and physical questions can or should be asked about a system. In the first case, the protein rapidly folds and is essentially in equilibrium between the folded and unfolded states. As a result, knowing the thermodynamic stability (ΔG), how that stability changes with mutation or changing environmental conditions, and how that protein refolds from a fully unfolded state informs its cellular behavior. In the second case, the protein will likely only fold once, and that folding process is slow enough that it will likely interact extensively with cellular folding machinery, possibly during synthesis by the ribosome. Additionally, *in vivo* the second, slower unfolding protein is unlikely to be in equilibrium with the unfolded state, and is instead kinetically trapped in the native state. As a result, the thermodynamic stability is less relevant to the cellular behavior than the rate of unfolding.

1.3.3 Timescales accessible by biophysical experiments

The repertoire of biophysical methods, each with its own unique set of technical restrictions and accessible timescales, is too vast to explore here. Covalent labeling methods alone have so many various chemistries and instrumentations that dissecting each of them individually is a daunting task. This thesis focuses on two covalent labeling methods, HDX-MS and HRF-MS. As explained above, when labeling experiments are in an EX1 regime the observed rate reports on the kinetic barrier. To access this regime k_{close} must be much slower than $k_{intrinsic}$, so the intrinsic rates provide a ceiling for which timescales can be accessed. At physiological pH HDX intrinsic rates are one the order of 10-100 milliseconds. In EX2 conditions, where k_{close} must be much faster than $k_{intrinsic}$, the resulting data informs on the conformational ensemble, as explained above. This does mean that one cannot exactly measure the timescale of conformational transitions within that ensemble, but one can conclude that any conformational dynamics are much faster than the intrinsic rates under those solution conditions. Pulsed-labeling experiments are ideal for monitoring the kinetics of conformational transitions. The timescales accessible for pulsed-labeling experiments are limited by the lengths of pulses necessary for sufficient chemical

labeling. Pulse lengths for both x-ray and hydrogen peroxide based HRF are on the order of 10 microseconds. Pulse lengths for HDX depend on pH and the degree of protection of the conformations being studied. By manipulating the pH of the pulse, HDX pulsed labeling has been effectively used to study protein folding in the 100 millisecond to seconds regime, conformational transitions in viral fusion proteins on the seconds to minutes regime.[40,49,107]

1.4 Summary

The work in this thesis uses covalent labeling methods to investigate questions related to protein energy landscapes. In Chapter 2, x-ray hydroxyl radical footprinting/mass spectrometry (HRF-MS) is used to monitor the unfolding of proteins in denaturant and extract thermodynamic parameters related to stability. This work demonstrates the utility of HRF-MS in measuring protein stability with minimal sample, in complex solution conditions, and for proteins with non-two-state behavior. This work will be submitted for publication with me as the first author. In Chapter 3, hydrogen/deuterium exchange-mass spectrometry (HDX-MS) is used to probe the conformational ensemble of the SARS-CoV-2 spike protein. These experiments revealed a highly populated, previously uncharacterized state of the SARS-CoV-2 spike protein, with highly accessible receptor binding domains and exposed conserved regions of the trimeric interface. This work was published in *Nature Structural and Molecular Biology* with me as a co-first author.[107] In Chapter 4, multiple covalent labeling methods including HDX-MS, thiol labeling, and covalent ligand binding specific to the protein HaloTag to investigate the folding pathway of HaloTag *in vitro* and during translation by the ribosome. This work revealed that HaloTag populates an aggregation-prone intermediate during re-folding from denaturant, and that this intermediate is not populated during co-translational folding, preventing aggregation and improving the folding yield. This work was published in *Science Advances* with me as the third author.[83] In Chapter 5 I conclude by summarizing the future directions for the projects in this thesis.

1.5 References

1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

2. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

3. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).

4. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

5. Keedy, D. A. *et al.* Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *Elife* **4**, (2015).

6. Tollinger, M., Crowhurst, K. A., Kay, L. E. & Forman-Kay, J. D. Site-specific contributions

to the pH dependence of protein stability. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4545–4550 (2003).

7. Royer, C. A. Why and How Does Pressure Unfold Proteins? in *High Pressure Bioscience: Basic Concepts, Applications and Frontiers* (eds. Akasaka, K. & Matsuki, H.) 59–71 (Springer Netherlands, 2015).

8. Wallace, E. W. J. *et al.* Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell* **162**, 1286–1298 (2015).

9. Cox, C. D., Bavi, N. & Martinac, B. Biophysical Principles of Ion-Channel-Mediated Mechanosensory Transduction. *Cell Rep.* **29**, 1–12 (2019).

10. Kim, I. S. *et al.* Mechanism of membrane fusion induced by vesicular stomatitis virus G protein. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E28–E36 (2017).

11. Patel, A. *et al.* ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017).

12. Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S. & Pielak, G. J. Macromolecular crowding and protein stability. *J. Am. Chem. Soc.* **134**, 16614–16618 (2012).

13. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574–581 (2009).

14. Park, C., Zhou, S., Gilmore, J. & Marqusee, S. Energetics-based protein profiling on a proteomic scale: identification of proteins resistant to proteolysis. *J. Mol. Biol.* **368**, 1426–1437 (2007).

15. Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* **415**, 343–346 (2002).

16. Xia, K. *et al.* Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17329–17334 (2007).

17. Young, R. & Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem. J* **160**, 185–194 (1976).

18. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **16**, 589–597 (2009).

19. Braselmann, E., Chaney, J. L. & Clark, P. L. Folding the proteome. *Trends Biochem. Sci.* **38**, 337–344 (2013).

20. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & John R Yates, I. I. I. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).

21. Catherman, A. D., Skinner, O. S. & Kelleher, N. L. Top Down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* **445**, 683–693 (2014).

22. Karch, K. R. *et al.* Hydrogen-Deuterium Exchange Coupled to Top- and Middle-Down

Mass Spectrometry Reveals Histone Tail Dynamics before and after Nucleosome Assembly. *Structure* **26**, 1651–1663.e3 (2018).

23. Hvidt, A., Johansen, G., Linderstrøm Lang, K. & Vaslow, F. Exchange of deuterium and 18O between water and other substances. I. Methods. *C. R. Trav. Lab. Carlsberg Chim.* **29**, 129–157 (1954).

24. Krause, I. M. & Lindstrom-Lang, K. Exchange of deuterium and 18O between water and other substances. II. Alternative-methods. *C. R. Trav. Lab. Carlsberg Chim.* **29**, 367–384 (1955).

25. Hvidt, A. & Linderstrom-Lang, K. Exchange of deuterium and 18O between water and other substances. III. Deuterium exchange of short peptides, Sanger's A-chain and insulin. *C. R. Trav. Lab. Carlsberg Chim.* **29**, 385–402 (1955).

26. Hvidt, A. & Linderstrom-Lang, K. The kinetics of the deuterium exchange of insulin with D2O; an amendment. *Biochim. Biophys. Acta* **16**, 168–169 (1955).

27. Hvidt, A. & Linderstrøm-Lang, K. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochim. Biophys. Acta* **14**, 574–575 (1954).

28. Hvidt, A. & Nielsen, S. O. Hydrogen Exchange in Proteins. in *Advances in Protein Chemistry* (eds. Anfinsen, C. B., Anson, M. L., Edsall, J. T. & Richards, F. M.) vol. 21 287–386 (Academic Press, 1966).

29. Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. Hydrogen exchange: the modern legacy of Linderstrøm-Lang. *Protein Sci.* **6**, 1101–1109 (1997).

30. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192–197 (1995).

31. Raschke, T. M. & Marqusee, S. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat. Struct. Biol.* **4**, 298–304 (1997).

32. Arrington, C. B. & Robertson, A. D. Kinetics and thermodynamics of conformational equilibria in native proteins by hydrogen exchange. in *Methods in Enzymology* vol. 323 104–124 (Academic Press, 2000).

33. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **25**, 158–170 (2006).

34. Ferraro, D. M., Lazo, N. D. & Robertson, A. D. EX1 hydrogen exchange and protein folding. *Biochemistry* **43**, 587–594 (2004).

35. Sivaraman, T. & Robertson, A. D. Kinetics of Conformational Fluctuations by EX1 Hydrogen Exchange in Native Proteins. in *Protein Structure, Stability, and Folding* (ed. Murphy, K. P.) 193–214 (Humana Press, 2001).

36. Zheng, J., Strutzenberg, T., Pascal, B. D. & Griffin, P. R. Protein dynamics and

conformational changes explored by hydrogen/deuterium exchange mass spectrometry. *Curr. Opin. Struct. Biol.* **58**, 305–313 (2019).

37. Morgan, C. R. & Engen, J. R. Investigating solution-phase protein structure and dynamics by hydrogen exchange mass spectrometry. *Curr. Protoc. Protein Sci.* **Chapter 17**, Unit 17.6.1–17 (2009).

38. Zhang, Y., Rempel, D. L. & Gross, M. L. Hydrogen Exchange Mass Spectrometry for the Analysis of Ligand Binding and Protein Aggregation. in *Hydrogen Exchange Mass Spectrometry of Proteins* 185–207 (John Wiley & Sons, Ltd, 2016).

39. Benhaim, M., Lee, K. K. & Guttman, M. Tracking Higher Order Protein Structure by Hydrogen-Deuterium Exchange Mass Spectrometry. *Protein Pept. Lett.* **26**, 16–26 (2019).

40. Hu, W. *et al.* Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7684–7689 (2013).

41. Hoang, L., Bedard, S., Krishna, M. M. G., Lin, Y. & Englander, S. W. Cytochrome c folding pathway: kinetic native-state hydrogen exchange. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12173–12178 (2002).

42. Maity, H., Maity, M. & Englander, S. W. How cytochrome c folds, and why: submolecular foldon units and their stepwise sequential stabilization. *J. Mol. Biol.* **343**, 223–233 (2004).

43. Walters, B. T., Mayne, L., Hinshaw, J. R., Sosnick, T. R. & Englander, S. W. Folding of a large protein at high structural resolution. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18898–18903 (2013).

44. Lim, S. A., Bolin, E. R. & Marqusee, S. Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange. *Elife* **7**, 3754 (2018).

45. Chen, J., Rempel, D. L., Gau, B. C. & Gross, M. L. Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. *J. Am. Chem. Soc.* **134**, 18724–18731 (2012).

46. Jha, S. K., Dasgupta, A., Malhotra, P. & Udgaonkar, J. B. Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry. *Biochemistry* **50**, 3062–3074 (2011).

47. Jha, S. K. & Udgaonkar, J. B. Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry*. *J. Biol. Chem.* **282**, 37479–37491 (2007).

48. Aghera, N. & Udgaonkar, J. B. Stepwise Assembly of β-Sheet Structure during the Folding of an SH3 Domain Revealed by a Pulsed Hydrogen Exchange Mass Spectrometry Study. *Biochemistry* **56**, 3754–3769 (2017).

49. Benhaim, M. A., Mangala Prasad, V., Garcia, N. K., Guttman, M. & Lee, K. K. Structural monitoring of a transient intermediate in the hemagglutinin fusion machinery on influenza

virions. *Sci. Adv.* **6**, eaaz8822 (05/2020).

50. Molday, R. S., Englander, S. W. & Kallen, R. G. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150–158 (1972).

51. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).

52. Connelly, G. P., Bai, Y., Jeng, M. F. & Englander, S. W. Isotope effects in peptide group hydrogen exchange. *Proteins* **17**, 87–92 (1993).

53. Nguyen, D., Mayne, L., Phillips, M. C. & Walter Englander, S. Reference Parameters for Protein Hydrogen Exchange Rates. *J. Am. Soc. Mass Spectrom.* **29**, 1936–1939 (2018).

54. Dobson, C. M., Lian, L.-Y., Redfield, C. & Topping, K. D. Measurement of hydrogen exchange rates using 2D NMR spectroscopy. *J. Magn. Reson.* **69**, 201–209 (1986).

55. Marion, D., Ikura, M., Tschudin, R. & Bax, A. Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399 (1989).

56. Kan, Z.-Y., Walters, B. T., Mayne, L. & Englander, S. W. Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16438–16443 (2013).

57. Kan, Z.-Y., Mayne, L., Chetty, P. S. & Englander, S. W. ExMS: data analysis for HX-MS experiments. *J. Am. Soc. Mass Spectrom.* **22**, 1906–1915 (2011).

58. Mayne, L. *et al.* Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation-mass spectrometry method. *J. Am. Soc. Mass Spectrom.* **22**, 1898–1905 (2011).

59. Pan, J., Zhang, S., Parker, C. E. & Borchers, C. H. Subzero Temperature Chromatography and Top-Down Mass Spectrometry for Protein Higher-Order Structure Characterization: Method Validation and Application to Therapeutic Antibodies. *J. Am. Chem. Soc.* **136**, 13065–13071 (2014).

60. Xu, G. & Chance, M. R. Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Anal. Chem.* **77**, 4549–4555 (2005).

61. Xu, G. & Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**, 3514–3543 (2007).

62. Sharp, J. S., Becker, J. M. & Hettich, R. L. Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry. *Anal. Chem.* **76**, 672–683 (2004).

63. Chance, M. R., Farquhar, E. R., Yang, S., Lodowski, D. T. & Kiselar, J. Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution. *J. Mol. Biol.* **432**, 2973–2984 (2020).

64. Takamoto, K. & Chance, M. R. Radiolytic protein footprinting with mass spectrometry to

probe the structure of macromolecular complexes. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 251–276 (2006).

65. Gupta, S., Feng, J., Chance, M. & Ralston, C. Recent Advances and Applications in Synchrotron X-Ray Protein Footprinting for Protein Structure and Dynamics Elucidation. *Protein Pept. Lett.* **23**, 309–322 (2016).

66. Chance, M. R. *et al.* A new method for examining the dynamics of macromolecules: Time-resolved synchrotron x-ray 'footprinting'. *Synchrotron Radiat. News* **11**, 7–16 (2008).

67. Gupta, S., Celestre, R., Feng, J. & Ralston, C. Advancements and Application of Microsecond Synchrotron X-ray Footprinting at the Advanced Light Source. *Synchrotron Radiat. News* **29**, 39–44 (2016).

68. Gupta, S., Celestre, R., Petzold, C. J., Chance, M. R. & Ralston, C. Development of a microsecond X-ray protein footprinting facility at the Advanced Light Source. *J. Synchrotron Radiat.* **21**, 690–699 (2014).

69. Zhang, B., Cheng, M., Rempel, D. & Gross, M. L. Implementing Fast Photochemical Oxidation of Proteins (FPOP) as a Footprinting Approach to Solve Diverse Problems in Structural Biology. *Methods* **144**, 94–103 (2018).

70. Gau, B. C., Sharp, J. S., Rempel, D. L. & Gross, M. L. Fast photochemical oxidation of protein footprints faster than protein unfolding. *Anal. Chem.* **81**, 6563–6571 (2009).

71. Espino, J. A. & Jones, L. M. Illuminating Biological Interactions with in Vivo Protein Footprinting. *Anal. Chem.* **91**, 6577–6584 (2019).

72. Chea, E. E. & Jones, L. M. Analyzing the structure of macromolecules in their native cellular environment using hydroxyl radical footprinting. *Analyst* **143**, 798–807 (2018).

73. Espino, J. A., Mali, V. S. & Jones, L. M. In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Anal. Chem.* **87**, 7971–7978 (2015).

74. Rinas, A., Mali, V. S., Espino, J. A. & Jones, L. M. Development of a Microflow System for In-Cell Footprinting Coupled with Mass Spectrometry. *Anal. Chem.* **88**, 10052–10058 (2016).

75. Kaur, U. *et al.* Proteome-Wide Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic Scale. *J. Proteome Res.* **17**, 3614–3627 (2018).

76. Sharp, J. S. *et al.* Flash Oxidation (FOX) System: A Novel Laser-Free Fast Photochemical Oxidation Protein Footprinting Platform. *J. Am. Soc. Mass Spectrom.* (2021) doi:10.1021/jasms.0c00471.

77. Weinberger, S. R., Chea, E. E., Sharp, J. S. & Misra, S. K. Laser-free Hydroxyl Radical Protein Footprinting to Perform Higher Order Structural Analysis of Proteins. *J. Vis. Exp.* (2021) doi:10.3791/61861.

78. Huang, W., Ravikumar, K. M., Chance, M. R. & Yang, S. Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: a protection factor analysis. *Biophys. J.* **108**, 107–115 (2015).

79. Scinto, S. L. *et al.* Bioorthogonal chemistry. *Nat Rev Methods Primers* **1**, (2021).

80. Los, G. V. *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).

81. Mollwitz, B. *et al.* Directed evolution of the suicide protein O^6-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* **51**, 986–994 (2012).

82. Gautier, A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15**, 128–136 (2008).

83. Samelson, A. J. *et al.* Kinetic and structural comparison of a protein's cotranslational folding and refolding pathways. *Sci Adv* **4**, eaas9098 (2018).

84. Capece, M. C., Kornberg, G. L., Petrov, A. & Puglisi, J. D. A simple real-time assay for in vitro translation. *RNA* **21**, 296–305 (2015).

85. Berg, H. C. *Random Walks in Biology*. (Princeton University Press, 1993).

86. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1309–1314 (1961).

87. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).

88. Levinthal, C. Are there pathways for protein folding? *Journal de Chimie Physique* **65**, 44–45 (1968).

89. Röder, K., Joseph, J. A., Husic, B. E. & Wales, D. J. Energy Landscapes for Proteins: From Single Funnels to Multifunctional Systems. *Advanced Theory and Simulations* **45**, 1800175 (2019).

90. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).

91. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).

92. Schwarz, G. & Seelig, J. Kinetic properties and the electric field effect of the helix-coil transition of poly(gamma-benzyl L-glutamate) determined from dielectric relaxation measurements. *Biopolymers* **6**, 1263–1277 (1968).

93. Wada, A., Tanaka, T. & Kihara, H. Dielectric dispersion of the -helix at the transition region to random coil. *Biopolymers* **11**, 587–605 (1972).

94. Du, D., Zhu, Y., Huang, C.-Y. & Gai, F. Understanding the key factors that control the rate of β-hairpin folding. *Proceedings of the National Academy of Sciences* **101**, 15915–15920 (2004).

95. Muñoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. Folding dynamics and mechanism of beta-hairpin formation. *Nature* **390**, 196–199 (1997).

96. Sadqi, M., Lapidus, L. J. & Muñoz, V. How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12117–12122 (2003).

97. Maxwell, K. L. *et al.* Protein folding: defining a standard set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**, 602–616 (2005).

98. Shamir, M., Bar-On, Y., Phillips, R. & Milo, R. SnapShot: Timescales in Cell Biology. *Cell* **164**, 1302–1302.e1 (2016).

99. Maillard, R. A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459–469 (2011).

100. Samelson, A. J., Jensen, M. K., Soto, R. A., Cate, J. H. D. & Marqusee, S. Quantitative determination of ribosome nascent chain stability. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13402–13407 (2016).

101. Jensen, M. K., Samelson, A. J., Steward, A., Clarke, J. & Marqusee, S. The folding and unfolding behavior of ribonuclease H on the ribosome. *J. Biol. Chem.* **295**, 11410–11417 (2020).

102. Carroll, E. C., Greene, E. R., Martin, A. & Marqusee, S. Ubiquitination modulates a protein energy landscape site-specifically with consequences for proteasomal degradation. Preprint at https://doi.org/10.1101/843631.

103. Waudby, C. A., Burridge, C., Cabrita, L. D. & Christodoulou, J. Thermodynamics of co-translational folding and ribosome-nascent chain interactions. *Curr. Opin. Struct. Biol.* **74**, 102357 (2022).

104. Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).

105. Cassaignau, A. M. E. *et al.* Interactions between nascent proteins and the ribosome surface inhibit co-translational folding. *Nat. Chem.* **13**, 1214–1220 (2021).

106. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13004–13009 (2006).

107. Costello, S. M. *et al.* The SARS-CoV-2 spike reversibly samples an open-trimer conformation exposing novel epitopes. *Nat. Struct. Mol. Biol.* **29**, 229–238 (2022).

Chapter 2

Monitoring local protein denaturation in complex environments with HRF-MS

The work presented here was led by me and done in close collaboration with Sayan Gupta and Corie Y. Ralston at Lawrence Berkeley National Laboratory and the Advanced Light Source. My contributions to this work include the conceptualization of the project, performance of experimental work, data analysis, and interpretation.

2.1 Abstract

Protein stability is a major determinant of cellular function, and improving our understanding of protein stability will greatly aid efforts in protein engineering and design. Despite its importance, traditional methods for measuring protein stability suffer from many limitations, including requiring significant quantities of highly purified protein under well-defined conditions. In this work, I develop methods to measure protein stability by combining X-ray hydroxyl radical footprinting (HRF) monitored by mass spectrometry (MS) with chemical denaturation. This approach can accurately determine the thermodynamic stability of a protein using several orders of magnitude less protein than the more traditional approaches such as circular dichroism and intrinsic fluorescence. Moreover, because both the footprinting and the MS detection can be carried out in complex environments, this approach is capable of measuring the stability of proteins in complex mixtures of biomolecules. Lastly, by monitoring the footprinting at the level of individual peptides within a protein, HRF-MS allows for the determination of stability for individual regions of a protein. Such local stability measurements can reveal subtle deviations from two-state behavior not detectable by more standard approaches. Overall, this approach allows for the accurate measurement of protein stability, and the unique strengths of this method extend our ability to measure stability to a much larger fraction of proteins.

2.2 Introduction

Proteins are large, complex, and essential macromolecules that perform a wide range of biological functions, including catalyzing chemical reactions, transporting molecules, and responding to external stimuli. To accomplish this, most proteins must fold into a structured ‘native’ conformation. It is well-known that this structural information is encoded in a protein’s amino-acid sequence.[1–3] Recent advances in machine learning algorithms have revolutionized the field of protein structure prediction, allowing for the fast and accurate prediction of such native state structures given only the amino acid sequence.[4,5] However, to accomplish their biological function, proteins need to access more than this folded native structure; proteins are dynamic, sampling alternative conformations with defined populations and rates of interconversion, all of which is commonly referred to as a protein’s conformational energy landscape.[6] Such energy landscapes are critical determinants of enzyme catalysis, allostery, signaling, misfolding, and

turnover among many other functions.[6,7] So while these revolutionary algorithms have solved the structure prediction problem, they are as of yet unable to describe this protein heterogeneity. Consequently, a major unsolved question is how protein sequence, along with post-translational modifications (PTMs) and interactions with other biomolecules, affect a protein's energy landscapes, rather than just its static ground-state structure.

The most basic description of a protein's energy landscape is its global or thermodynamic stability. Protein thermodynamic stability, $\Delta G_{folding}$, is defined as the difference in free energy between the folded state and the unfolded state. This energy difference dictates the relative population of each state ($\Delta G^{\circ}_{folding} = -RT\ ln\left(\frac{[Native]}{[Unfolded]}\right)$). Thus, stability is typically determined by measuring the population of each state under equilibrium conditions. Since most globular proteins are at least -5 kcal/mol under native conditions, it is impossible to measure these populations under native conditions and, therefore, their stability is determined by denaturing the protein, using chemical denaturants or temperature, monitoring the change in populations over the unfolding transition, and then extrapolating to physiological, or native conditions.[8] Importantly, this assumes that a protein can only occupy two states, the native state and unfolded state. As protein folding is a highly cooperative process this is often a valid assumption. However, for larger proteins, and for proteins with multiple domains, partially folded states are often accessible, requiring more complex models to extract stability parameters. A protein's global stability is a crucial feature of its function as, for most proteins, it determines the population of molecules in the functional native state. Similarly, stability determines the population of unfolded molecules, which can determine the propensity for misfolding or proteolytic degradation. Therefore, even when destabilizing mutations have minor effects on native state structure, catalysis, or binding, they can have large effects on steady-state protein levels.[9–11]

Traditional methods for measuring protein stability rely on optical probes of structure, such as circular dichroism (CD) or intrinsic fluorescence, which require large amounts (usually milligrams) of highly purified, isolated proteins. This limits their application to abundant proteins that can be easily generated, usually via overexpression in E. coli and isolated, or purified. Because all biomolecules in solution contribute to the detected signal, such stability measurements cannot be made in the presence of other biomolecules, such as seen in complex biological environments, limiting most measurements of protein stability to simple model systems under idealized conditions.

Recently, several methods have been developed that overcome some of these limitations, such as pulse proteolysis, SUPREX, and SPROX.[12–17] All rely on quantitively determining the population of unfolded molecules by selective modification. When coupled with mass spectrometry, they allow the analysis using small quantities of protein, as well as the ability to monitor proteins in complex mixtures. While extremely powerful, these modification approaches also have limitations. For instance, all rely on a short pulse, ranging from several seconds to several minutes, during which the protein is selectively modified. In order to accurately detect different populations, the protein conformations cannot interconvert during this pulse. However, the

timescale of the pulse is often longer than the timescale of folding and unfolding for many proteins, and thus cannot be used on a large fraction of the proteome.[18] Second, these methods also make use of a two-state assumption to measure global stability, but, again, not all proteins fold via a two-state mechanism (populating only a native and unfolded conformation). For proteins that populate intermediates, or for proteins with multiple domains capable of independent folding, a method that provides multiple local probes for folding is needed.

In this chapter, I develop and utilize another selective modification approach, X-ray-based hydroxyl radical footprinting (HRF), which again, monitors the relative population of unfolded molecules as a function of chemical denaturant. X-ray-based HRF is a powerful tool that uses very short exposures of synchrotron-generated X-ray beams to produce hydroxyl radicals from water molecules in solution.[19,20] The resulting radicals rapidly and irreversibly modify nearby proteins residues resulting in oxidized sidechains. There are multiple possible modifications, the most common being the net addition of oxygen with a change in mass (+16 Da) that can be detected by mass spectrometry. As the radicals are generated from solvent and are water soluble, the observed reactivity depends on the solvent accessibility of that side chain. Therefore, by monitoring the extent of modification for each side chain under different conditions, information about changes in protein structure can be obtained. In many ways, the approach outlined in this work is similar to SPROX, which monitors the modification of a single amino acid type, methionine, by hydrogen peroxide.[16,17] However, the faster labeling kinetics of HRF allows for the use of much shorter labeling times, increasing the number of proteins amenable to this method. Additionally, because most amino acids are reactive, the modification sites are distributed throughout the whole protein, which allows one to monitor the unfolding of individual peptides within a protein, providing local information relating to stability. Thus, this HRF approach can measure the energetics of large complex proteins in complex mixtures at the structural level of individual peptides.

2.3 Results

2.3.1 Protein denaturation monitored by HRF-MS accurately determines the global stability of the model protein ecRNH*

To determine whether HRF-MS coupled with chemical denaturation can accurately measure the global stability of a protein, I turned to the well-characterized single-domain protein, E. coli RNase H (ecRNH*, where the * indicates all cysteines have been mutated to alanine), whose equilibrium stability has been measured by many techniques.[21–24] Samples of protein equilibrated in chemical denaturant (urea) were subjected to a short pulse of X-rays and then evaluated by mass spectrometry for their relative amounts of modification. HRF was performed on each sample by flowing through a capillary tube perpendicular to a synchrotron beam such that the duration of exposure (pulse length) was either 500 or 1000 microseconds and frozen for later analysis (see Methods). The degree of oxidation in each sample was then measured by mass spectrometry. Fig. 2-1A demonstrates that under conditions that populate the unfolded state (high urea), MS reveals much greater oxidation than under native conditions, where the native state predominates, reflecting exposure of reactive side chains upon denaturation. Across all

denaturant conditions, the overall mass distributions appear to be a sum of this low reactivity population corresponding to the folded state and the high reactivity population, corresponding to the unfolded state, as expected for a cooperative process. When these mass spectra are quantified by determining the weighted average of the mass distributions, a clear cooperative transition is observed (Fig. 2-1B). These data can be fit using a traditional two-state linear-extrapolation model (LEM, see methods), to obtain free energies of folding ($\Delta G_{folding}$) and m-values that are in agreement with the stability determined by more traditional methods such as CD. (Fig. 2-1C, Table 2-1). Thus, HRF-MS can be used to faithfully determine global stability in a simple two-state model system.

One concern with HRF is the potential for the oxidative modifications to destabilize the protein and induce unfolding, which would result in lower apparent stability. This can be combatted with the shortest possible pulse duration, reducing the probability of unfolding during the timescale of the pulse. The agreement between the melts obtained from two different pulse lengths supports the assumption that modifications are not inducing unfolding on the timescale of the experiment. Additionally, the agreement between HRF and CD further supports the validity of HRF-MS as a method for monitoring denaturation. Based on well-characterized two-state folding proteins, the majority of proteins will not significantly re-equilibrate during the one-millisecond pulse used in HRF.

Of note, this approach used only 0.1 micrograms of protein per denaturant condition, allowing for the determination of stability using only 4 micrograms of protein. This is in comparison to traditional optical methods, where one usually uses ~50 micrograms per denaturant condition, providing a savings of ~500-fold.

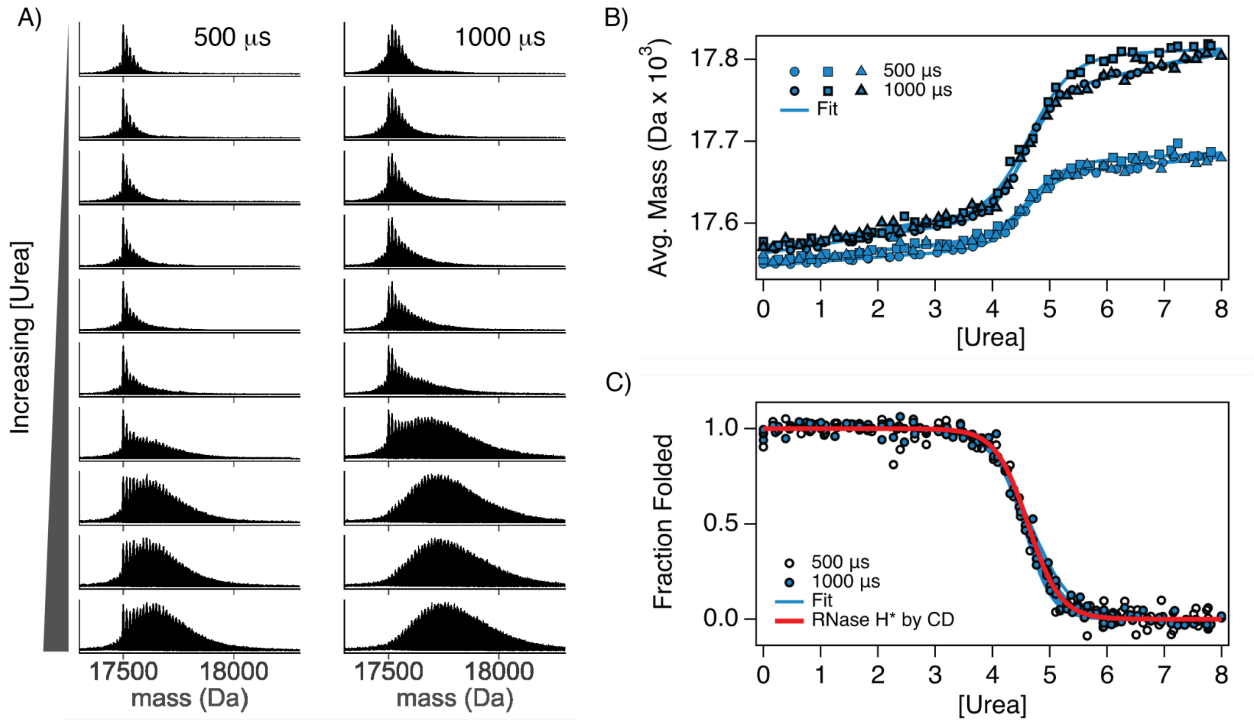


Figure 2-1. Global denaturant melt of ecRNH* by HRF reproduces known stability. (A) Deconvoluted mass spectra for ecRNH* in varying concentrations of denaturant for two different exposure times, 500 and 1000 µs. (B) Weighted average mass for deconvoluted mass spectra as a function of denaturant for three replicates. (C) Overlay of triplicate denaturant melts, at two exposure times, normalized to fraction folded. The expected fraction folded based on denaturant melts monitored by circular dichroism is shown as a red line.

	m-value (kcal/(mol*M))	**C_m (M)**	**$\Delta G_{folding}$ (kcal/mol) (Fit m-value)**	**$\Delta G_{folding}$ (kcal/mol) (Calculated m-value)**
HRF-MS (500 µs)	2.1 ± 0.3	4.58 ± 0.05	9.7 ± 1.2	9.2 ± 0.1
HRF-MS (1000 µs)	2.0 ± 0.4	4.60 ± 0.06	9.1 ± 1.5	9.3 ± 0.1
CD	2.10 ± 0.09		9.7 ± 0.4	

Table 2-1. Thermodynamic parameters from ecRNH* HRF-MS denaturation. Shown are the averages ± standard deviations resulting from triplicate measurements. Parameters resulting from CD taken from Raschke et al.[25] ΔG were calculated as the product of the C_m and m-value (ΔG = –m-value*C_m) using either the measured m-value or the m-value calculated from the observed relationship between m-value and protein size (m-value = length*0.0013)[13,26]

2.3.2 Local measurements of denaturant induced modifications can be monitored using peptide HRF-MS

As demonstrated above, using HRF and monitoring the denaturant-induced modifications by intact MS returns accurate thermodynamic parameters. However, this approach relies on the ability to obtain intact mass spectra. For some proteins (especially very large proteins) this can be technically challenging. In addition, while intact measurements yield measurements of global stability, they do not report on subtle differences throughout the molecule, such as changes in local stability. Thus, I performed the same chemical denaturation HRF experiment, but this time analyzed at a more local level - I digested (using trypsin) and analyzed the oxidation susceptibility of individual peptides (see Methods). Of the eleven tryptic peptides detected for ecRNH*, nine showed cooperative transitions consistent with unfolding (Fig. 2-2A). Note that only regions that have a differential susceptibility to oxidation in the folded and unfolded state will produce a denaturation profile. These peptides provide local measurements of stability at the corresponding regions of the protein.

Analysis of these peptides proved to be more challenging than the intact protein. Although cooperative, the plots of peptide modification as a function of denaturation showed peptide-specific non-linear baselines. Therefore, instead of using a model with linear baselines, I used a modified LEM, with an exponential rather than a linear native baseline and a linear denatured baseline (the denatured baseline did not have enough points to support fitting with an exponential) to obtain denaturation profiles (the fraction folded as a function of denaturant) (see Methods). The resulting denaturation profiles overlay well with the global melt obtained by intact HRF/MS (Fig. 2-2B). For all peptides monitored in this way, this fitting approach returned denaturant concentration midpoints (C_m) within 0.2 M of the global intact melt (Fig. 2-2C, Table 2-2). The C_m determined for each peptide showed higher variation than the C_m determined for the intact protein and each fit had poorly determined m-values (Fig. 2-2D). This is perhaps not unexpected, since this modified model increases the total number of fit parameters from six to seven, increasing the risk of overfitting. In addition, not all peptides were amenable to this exponential baseline fit (for example see peptide 47-60). However, it is well known that m-values are not always needed and that the C_m is often a more reliable parameter to determine.[27] Moreover, m-values are known to correlate with protein size and therefore can be independently predicted based on the protein size.[13,26] Using the predicted ecRNH* m-value, this modified LEM approach returned a $\Delta G_{Folding}$ in agreement with the known value determined by CD (Table 2-2).

Figure 2.2C diagrams the C_m determined for each tryptic fragment in RNase H. In a perfectly two-state system, the C_m should be independent of the observation method (ie. CD, intact HRF-MS, or peptide HRF-MS). While all peptides report a C_m consistent with global unfolding, there appear to be some systematic deviations, which may suggest a deviation from two-state behavior (see next section). Future work will determine if the difference between peptide C_m and global C_m represents a genuine population of one or more partially folded conformations.

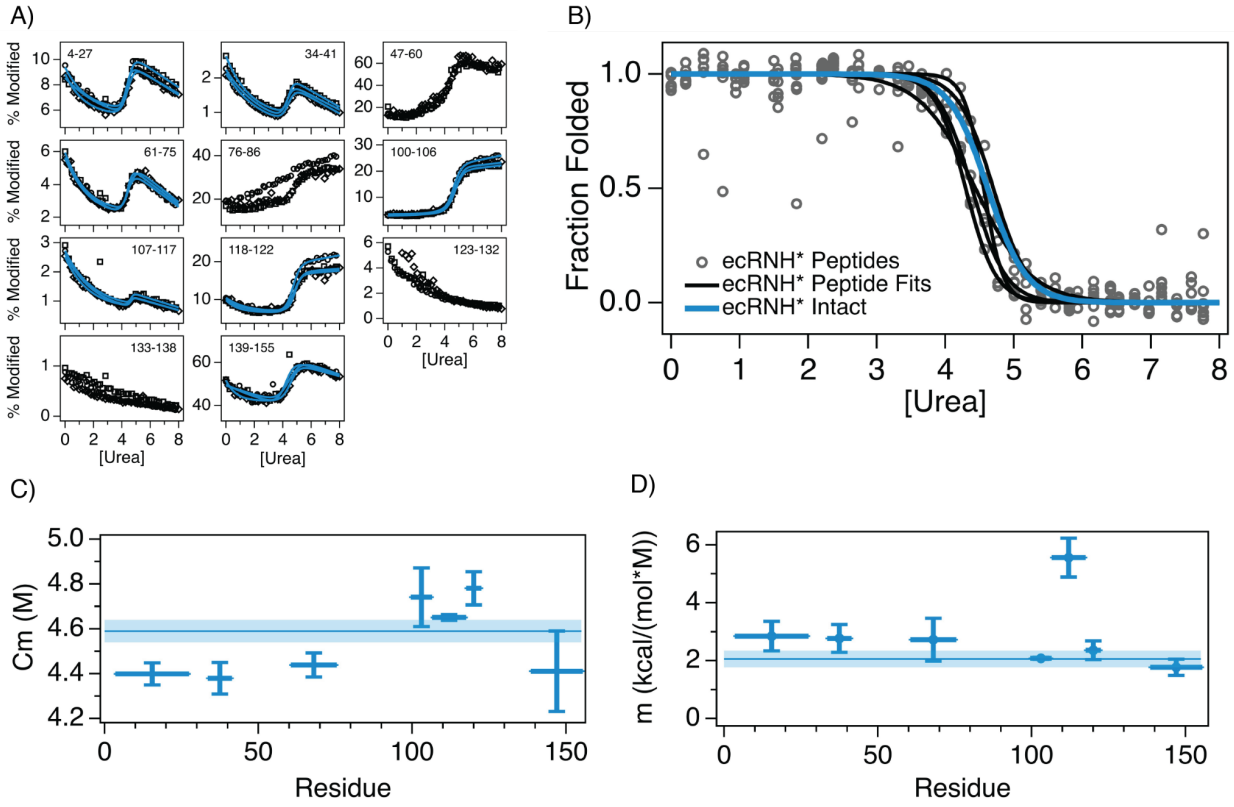


Figure 2-2. Denaturant melt of ecRNH* by HRF monitored at the peptide level. (A) Percent modified vs. denaturant concentration for eleven tryptic peptides from ecRNH*. Residue numbers are indicated in each panel. Blue lines represent fits to a modified LEM with an exponential native baseline. (B) Overlay of denaturant melts for all tryptic peptides with observable transitions, normalized to fraction folded. The expected fraction folded based on intact melts shown in figure 2-1 is shown as a blue line. (C) Concentration midpoints (C_m) and (D) m-values for each tryptic peptide. Sequence position for each peptide is represented by the x-axis. Error bars represent the standard deviation from triplicate measurements. Solid lines indicate values from intact melts, with shaded regions representing the standard deviation from triplicate measurements.

ecRNH* Peptide Stability	m-value (kcal/(mol*M))	C_m (M)	ΔG (kcal/mol) (Fit m-value)	ΔG (kcal/mol) (Calculated m-value)
Peptide 1 (4-27)	2.8 ± 0.5	4.40 ± 0.05	12.3 ± 1.9	8.86 ± 0.10
Peptide 2 (34-41)	2.7 ± 0.5	4.38 ± 0.07	12.0 ± 2.1	8.82 ± 0.14
Peptide 3 (61-75)	2.7 ± 0.7	4.44 ± 0.05	12.0 ± 3.2	8.95 ± 0.11
Peptide 4 (100-106)	1.98 ± 0.08	4.68 ± 0.06	9.3 ± 0.3	9.42 ± 0.11
Peptide 5 (107-117)	5.5 ± 0.7	4.62 ± 0.05	25.4 ± 3.4	9.31 ± 0.10

Peptide 6 (118-122)	2.3 ± 0.3	4.78 ± 0.07	11.2 ± 1.7	9.63 ± 0.15
Peptide 7 (139-155)	2.0 ± 0.2	4.57 ± 0.03	8.9 ± 0.9	9.21 ± 0.06
Peptide Average		4.55 ± 0.02		9.17 ± 0.10
CD	2.10 ± 0.09		9.7 ± 0.4	

Table 2-2. Thermodynamic parameters from ecRNH* HRF-MS denaturation monitored at the peptide level. Peptide sequence positions are noted in brackets. Shown are the averages ± standard deviations resulting from triplicate measurements. Parameters resulting from CD taken from Raschke et al.[25] ΔG were calculated as the product of the C_m and m-value ($\Delta G = -\text{m-value}*C_m$) using either the measured m-value or the m-value calculated from the observed relationship between m-value and protein size (m-value = length*0.0013)[13,26]

2.3.3 Monitoring the chemical-induced denaturation of multiple proteins simultaneously using HRF-MS

One weakness of traditional spectroscopic approaches is the need for purity. With a mixture of proteins, each will contribute to the observed signal. When samples contain multiple proteins, the observed results will produce either multiple transitions, which can be difficult to assign to individual components, or a single broadened transition which when erroneously treated as a single protein, returns thermodynamic values that are inconsistent with any of the proteins in solution. LC/MS-based approaches are ideally suited to handle this heterogeneity. Thus, HRF-MS should be amenable to any protein mixture as long as the proteins of interest (or peptides resulting from the digestion of those proteins) are separable by LC or have non-overlapping mass distributions. To demonstrate this, I performed chemical denaturation and HRF-MS on a mixture of two proteins, ecRNH* and T4 lysozyme (T4L*, where again the * represents a cysteine-free construct). ecRNH* and T4L* have non-overlapping mass distributions, allowing for both proteins to be independently monitored by intact HRF-MS, despite being a mixture (Fig. 2-3A).

Monitoring each mass distribution as a function of denaturant allows for the assignment of stability to each component (Fig. 2-3B). In either intact HRF-MS (Fig. 2-3C) or peptide HRF-MS (Fig. 2-3D), the signals from each protein are easily monitored independently, resulting in two distinct melts, from which thermodynamic values can be extracted that are consistent with the known stabilities for both ecRNH* and T4L* (Table 2-3). By comparison, CD measurements on the mixture do not generate independent stability measurements. When this same equimolar mixture of ecRNH* and T4L* is monitored by CD, a single broad transition is observed (Fig. 2-3C). This broad transition does not reflect the behavior of either component in the system, nor does it reflect any linear combination or average of the two. When fit using a traditional LEM, the resulting stability is a non-sensical $\Delta G_{folding}$ (6.9 kcal/mol), lower than either of the $\Delta G_{folding}$ for ecRNH* (9.06 kcal/mol) or T4L* (12.02 kcal/mol). Thus, HRF-MS provides a unique method for quantitatively determining the stability of individual components within a mixture. While in this case there are two different proteins, one could imagine a scenario where these instead are two domains within the same protein. In this hypothetical, it would be easy to assume the two-domain protein unfolded cooperatively and a two-state model would be chosen. However,

with the peptide HRF-MS approach, it would be apparent that each domain has a different unfolding transition and accurate thermodynamic parameters could be extracted. In future work, we will use this approach to dissect the energetics of multidomain proteins.

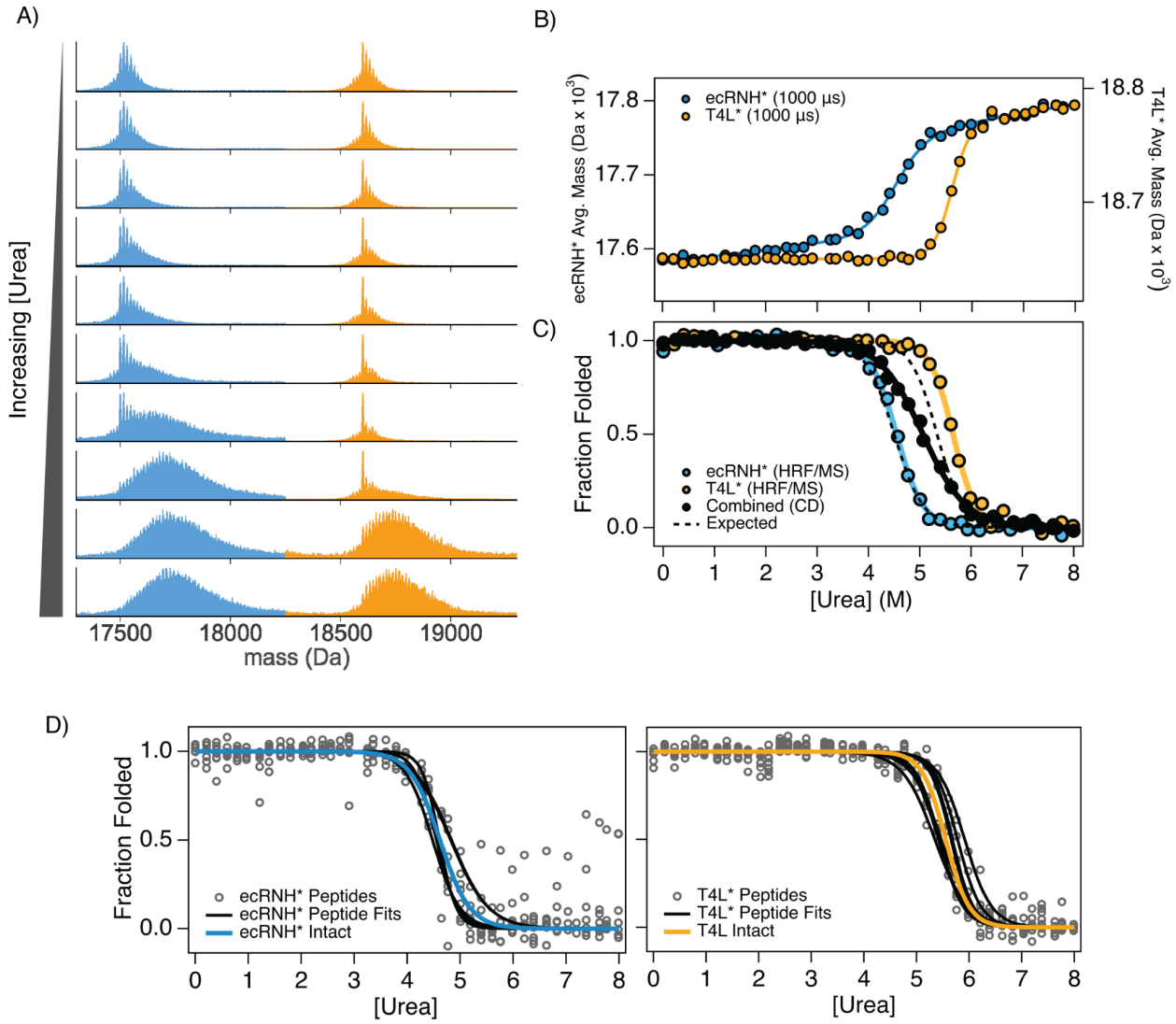


Figure 2-3. Simultaneous denaturant melt of ecRNH* and T4L*. (A) Deconvoluted mass spectra for a mixture of ecRNH* (blue) and T4L* (orange) in varying concentrations of denaturant for 1000 µs exposure time. Blue and orange data are normalized separately for ease of visualization. (B) Weighted average mass for deconvoluted mass spectra of ecRNH* (blue) and T4L* (orange) as a function of denaturant. (C) Overlay of denaturant melts normalized to fraction folded. Melts obtained by HRF are shown for ecRNH* (blue) and T4L*(orange). A melt of an equimolar mixture of ecRNH* and T4L* obtained by CD is shown as a black solid line. Expected fraction folded for isolated ecRNH* and T4L* based on denaturant melts monitored by CD are shown as dotted lines. Overlay of denaturant melts for all ecRNH* (D) and T4L* (E) tryptic peptides with observable transitions, normalized to fraction folded. The expected fraction folded based on intact melts is shown as blue (ecRNH*) and orange (T4L*) lines.

ecRNH*/T4L* Stability	**m-value (kcal/(mol*M))**	**C_m (M)**	**ΔG (kcal/mol) (Fit m-value)**	**ΔG (kcal/mol) (Calculated m-value)**
ecRNH* (isolated - CD)	2.05	4.50	9.23	9.07
T4L* (isolated - CD)	2.28	5.34	12.15	11.38
Combined (CD)	1.37	5.06	6.91	
ecRNH* (combined - HRF-MS)	1.94	4.50	8.72	9.06
T4L* (combined - HRF-MS)	2.67	5.64	15.03	12.02

Table 2-3. Thermodynamic parameters from the isolated and combined denaturation of ecRNH* and T4L*. Shown are the results from single experiments. ΔG were calculated as the product of the C_m and m-value (ΔG = –m-value*C_m) using either the measured m-value or the m-value calculated from the observed relationship between m-value and protein size (m-value = length*0.0013)[13,26]

.

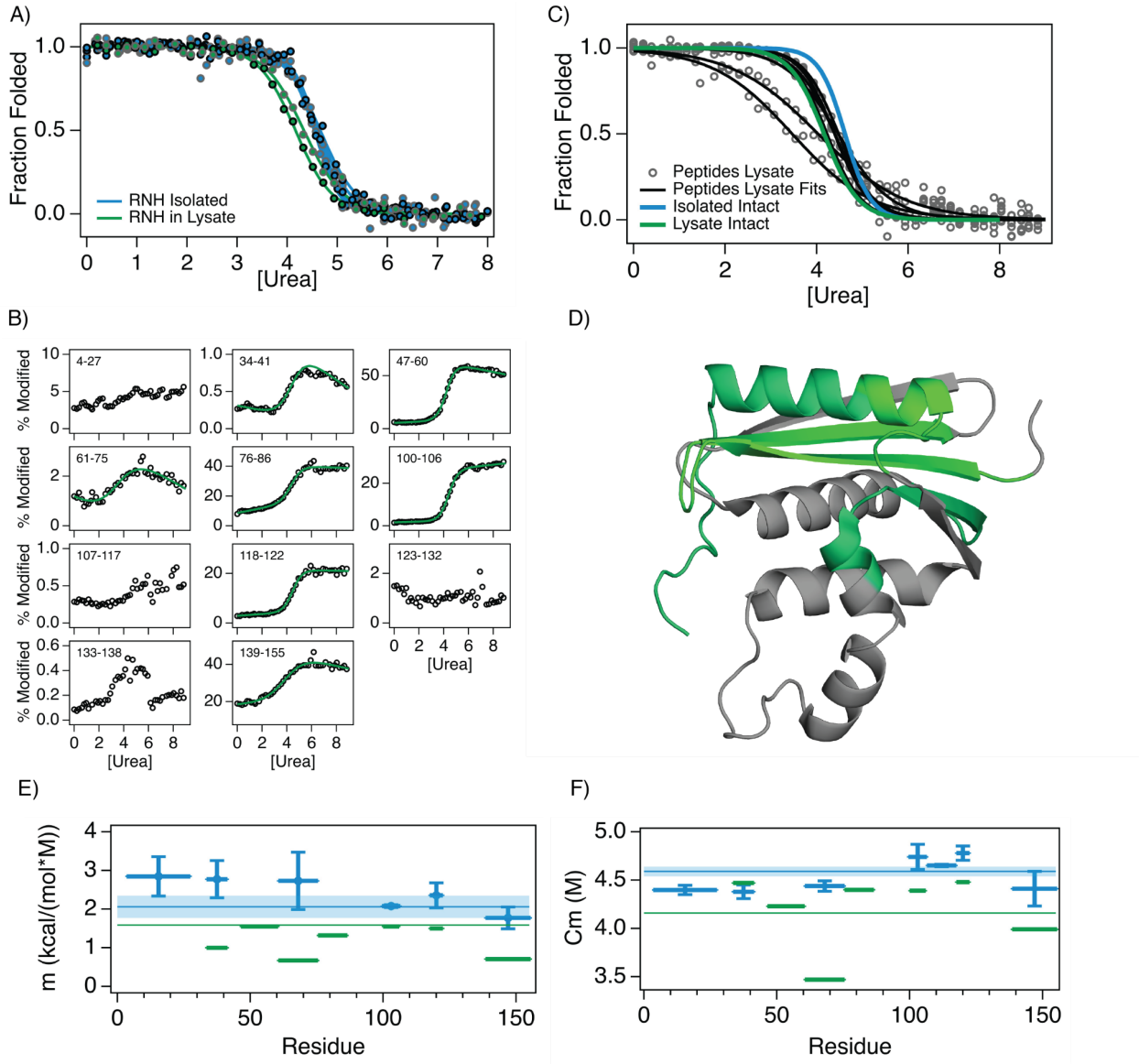


Figure 2-4. Denaturant melt of ecRNH* in the presence of E. coli lysate. (A) Overlay of denaturant melts normalized to fraction folded. Melts obtained by HRF are shown for isolated ecRNH* (blue) and ecRNH* in the presence of E. coli lysate (green) for two different exposure times, 500 us (gray outline) and 1000 us (black outline). (B) Percent modified vs. denaturant concentration for eleven tryptic peptides from ecRNH*. Residue numbers are indicated in each panel. Green lines represent fits to a LEM. (C) Overlay of denaturant melts for all ecRNH* tryptic peptides, in the presence of lysate, with observable transitions, normalized to fraction folded. The expected fraction folded based on intact melts shown in Figure 2-4A are shown as solid lines for isolated ecRNH* (blue) and ecRNH* in lysate (green). (D) Structure of ecRNH (PDB ID: 2RN2). Peptides with depressed m-values are shown as dark green. Peptides with observable transitions in isolation but no observable transition in lysate are shown as dark green (E) m-values and (F) concentration midpoints (C_m) for isolated ecRNH* (blue) and ecRNH* in the presence of lysate (green) tryptic peptides. The sequence position for each peptide is represented by the x-axis. Error bars represent the standard deviation from triplicate measurements. Solid lines indicate values from intact melts, with shaded regions representing the standard deviation from triplicate measurements.

2.3.4 Monitoring the stability of unpurified proteins using HRF-MS

The ability to perform melts in clarified lysates allows for bypassing purification steps. Additionally, the ability to perform denaturant melts in such complex environments allows for investigations of how these biological components affect protein stability in cells, experiments that are not amenable to traditional approaches. To further demonstrate the capability of this approach to deal with sample complexity, I performed a denaturant melt on ecRNH* in clarified E. coli lysate. The soluble components of E. coli were isolated, and diluted 10-fold into varying concentrations of denaturant with 1 μM ecRNH* added (see methods) and HRF-MS was performed.

Both the intact and peptide HRF-MS denaturation melts revealed unexpected differences that suggest perturbations to the energy landscape for ecRNH*. The intact HRF-MS melt of ecRNH* in lysate has a lower C_m and apparent m-value compared to ecRNH* in isolation (Fig 2-4A). A lower C_m suggests some destabilization of ecRNH*, and the lower m-value suggests either a change in the structure of the native state or a loss of two-state behavior. The peptide HRF-MS melts similarly show apparent two-state transitions (Fig 2-4B). However, unlike ecRNH* in isolation, not all peptide melts are consistent with the global transition (Fig 2-4C). Two peptides show depressed m-values (0.6 kcal/(mol*M), compared to 1.5 kcal/(mol*M) for ecRNH* in lysate and 2.1 kcal/(mol*M) for ecRNH* in isolation). Additionally, one peptide (residues 4-27) that has a clear unfolding transition for ecRNH* in isolation shows no transition in lysate. Disagreement between different probes of folding in a denaturant melt is strong evidence of a lack of two-state folding behavior.[8] However, unlike disagreement between spectroscopic probes, differences in peptide HRF-MS denaturation melts allows for structural interpretation. When mapped to the structure (Fig. 2-4D), perturbed peptides that have either a reduced m-value or no transition form the center of the beta-sheet and the c-terminal helix. Interestingly, these are the regions that are unfolded in the folding intermediate populated by ecRNH*, possibly suggesting that this folding intermediate is selectively stabilized by one or more components in the clarified E. coli lysate.[25,28,29] It is important to note that the buffer conditions between the isolated sample (20 mM NaOAc, 50 mM KCl, pH 5.5) and the lysate sample (PBS, pH 7.4) were not identical. More experiments will be needed to identify the precise cause and mechanism of this non-two-state behavior and destabilization.

2.4 Discussion

In this study, I demonstrated that HRF-MS, when combined with chemical denaturation, is a viable approach for measuring global and local protein thermodynamic stability. Additionally, I have demonstrated several advantages that HRF-MS has over both traditional methods and newer approaches for measuring thermodynamic stability. Specifically, HRF-MS requires less sample, is applicable to more systems, and has less stringent purity requirements. Compared to traditional methods like circular dichroism (CD) or intrinsic fluorescence, HRF-MS requires orders of magnitude less sample to measure stability. Compared to other recently developed methods such as pulse proteolysis, SUPREX, and SPROX the HRF-MS approach described here is amenable to a larger number of systems due to the faster labeling kinetics, broader reactivity, and less

stringent sample requirements. Our results also show that the less stringent purity requirements allow for HRF-MS to monitor specific proteins in mixtures, which can allow for bypassing labor-intensive purification steps, and opens up new possibilities for investigating the effects of biological components on protein stability.

In addition to the advantages outlined above, HRF-MS provides local information that can reveal more detailed energetics of more complex protein conformational landscapes. The ability to measure the stability of multiple proteins by monitoring their peptides is analogous to how we can also measure the stability of multiple domains within larger proteins. This could lead to more detailed descriptions of inter-domain energetic coupling. Additionally, the results for ecRNH* within lysate highlight how HRF-MS is capable of detecting subtle deviations from two-state behavior. In this specific case, further experiments are needed to identify the precise cause and mechanism of this behavior. However, these findings underscore the potential of HRF-MS to investigate the stability determinants of larger multi-domain proteins and non-two-state proteins.

This approach outlined in this work is likely to be generalizable and increasingly accessible. Generating a signal difference between folded and unfolded protein in HRF relies on buried hydrophobic residues in the native state, which is a nearly universal quality of globular proteins. This feature makes it more broadly applicable compared to other MS-based methods that rely instead on the selective aggregation of the unfolded state, like Thermal Proteome Profiling (TPP), or changes in protease susceptibility, like Limited Proteolysis (LiP).[30–37] Additionally, unlike TPP and LiP, HRF-MS combined with chemical denaturation provides rigorous thermodynamic information. However, TPP, LiP, and other methods have demonstrated the ability to probe many proteins in parallel by mass spectrometry, and as the approach described here is also an MS-based method, it can similarly take advantage of advances made by the proteomics field, further expanding the range of systems for which quantitative thermodynamic parameters can be measured. While HRF-MS is technically challenging, requiring a synchrotron and a mass spectrometer, mass spectrometers are increasingly available in core facilities, and bench-top hydroxyl radical footprinting instruments are now commercially available.[38] The accessibility of these instruments will facilitate the adoption of HRF-MS in a broader range of research settings.

In conclusion, the combination of HRF-MS with chemical denaturation represents a powerful tool for studying protein thermodynamic stability in complex biological systems. Its advantages over traditional methods and compatibility with a wide range of systems make it a valuable addition to the toolbox of techniques for investigating protein structure, stability, and function. Future work will focus on optimizing this method for high-throughput applications, expanding the range of proteins and biological systems that can be studied, and elucidating the molecular mechanisms underlying protein stability and folding dynamics in native-like conditions.

2.5 Methods

2.5.1 Protein expression and purification

Plasmids were transformed into Rosetta 2(DE3)pLysS cells for expression. One liter of cells were induced at OD = 0.6 with 1mM IPTG and grown at 37 °C for 3 hours. Cells were then pelleted, resuspended in 25 mM Tris-HCl, 10 mM NaCl, 0.1 mM EDTA pH 8.0, and lysed via sonication. Cell debris was then pelleted (ss34, 14k rpm, 30 minutes at 4 °C) and the resulting supernatant was filtered (0.22 um).

For the purification of ecRNH*, the clarified lysate was purified first over a HiTrap Heparin column (GE Healthcare) at pH 8. Peak fractions were pooled and diluted 3-fold with doubly deionized water. The pH of the solution was then adjusted to 5.5 using dilute HCl. The resulting aggregates were removed by another round of centrifugation and filtration. The filtered sample was purified over a HiTrap S column (GE Healthcare). Protein was then concentrated and buffer-exchanged into ammonium bicarbonate for subsequent freeze-drying and storage. Prior to use, the freeze-dried protein was resuspended in 20 mM NaOAc, 50 mM KCl, pH 5.5, and filtered (0.22 um).

For the purification of T4L*, clarified lysate was loaded onto a Capto S column, washed with 25 mM Tris, 10 mM NaCl, pH 8.0, and eluted with a gradient from 10 mM to 300 mM NaCl (25 mM Tris pH 8.0). Fractions containing T4L* were pooled, concentrated, and run on a S200 (GE Healthcare) column in PBS pH 7.4 (Sigma Aldrich P4417). Protein was stored at 4 °C until use.

2.5.2 Preparation of soluble E. coli lysate

E. coli K12 was grown overnight at 37 °C in 50 ml of LB media and harvested by centrifugation. The cell pellet was resuspended in 50 ml of 20 mM Tris–HCl, 250 mM NaCl, 10 mM EDTA pH 8.0 then pelleted again by centrifugation. The washed cell pellet was resuspended in 10 ml of 20 mM Tris–HCl, 50 mM NaCl, 1 mM EDTA pH 8.0 with added protease inhibitor and lysed by lysozyme treatment and repeated freeze-thaw cycles. Cell debris was then pelleted (ss34, 14k rpm, 30 minutes at 4 °C) and the resulting supernatant was filtered (0.22 um). Nucleic acids were digested by incubating supernatant with 0.1 mg/ml DNase I and 0.1 mg/ml RNase A. MgCl2 and CaCl2 were added to 2.5 mM and 1.0 mM, respectively, for this digestion reaction. To remove small metabolites and digested nucleic acids, the lysate was buffer exchanged using an amicon concentrator with a 3 kDa cutoff, first against 20 mM Tris–HCl, 250 mM NaCl, pH 8.0 and then against Phosphate Buffered Saline (PBS, Sigma Aldrich P4417). The lysate was sterilized by passing through a 0.22 μm syringe filter and stored at −20 °C until used.

2.5.3 X-ray hydroxyl radical footprinting

HRF experiments were carried out at the Advanced Light Source beamline 3.3.1 using a standard microfluidic set-up.[19,39,40] Protein concentrations were as follows: 5 uM for all isolated proteins, 2.5 uM each for the combined ecRNH* and T4L* experiments, and 1 uM ecRNH* in lysate.

ecRNH* and T4L* experiments were in 20 mM NaOAc, 50 mM KCl, pH 5.5. Lysate experiments were in PBS (Sigma Aldrich P4417). Samples were subject to X-ray exposures between 500 and 1250 us then immediately quenched with methionine amide to prevent secondary oxidations and stored at -80 °C until LC-MS analysis.

2.5.4 Intact protein mass spectrometry and analysis

Intact proteins were injected directly at a flow rate of 50 uL/min in 95% Buffer A (0.1% Formic Acid) 5% buffer B (100% Acetonitrile, 0.1% Formic Acid) onto a mAbPac column (Thermo Fisher) stored in a column oven maintained at 50 °C. Residual buffer and salts were removed by maintaining flow for 4 minutes with a valve diverting to waste. Proteins were then eluted with a gradient of 5-20% buffer B at a flow rate of 50 µL/min over 30 seconds, 20-70% buffer B over 4 minutes, then 70-90% over 30 seconds. The column was then subjected to a sawtooth wash and equilibrated at 5% buffer B prior to the next injection. Proteins were eluted directly into a Q Exactive Orbitrap Mass Spectrometer operating in positive mode (HESI source settings: Spray voltage 4.0 kV, Capillary temp. 325 C, S-lens RF level 70, Aux gas heater temp. 150 °C, In-source CID 15.0 eV — MS Settings: 6 microscans, resolution 140000, AGC target 5e6, maximum IT 100 ms, scan range 600-3000 m/z).

Intact mass spectra were summed across the entire chromatographic peak, then exported from Xcalibur and imported into Unidec.[41,42] Spectra were deconvoluted with Unidec, to remove charge state redundancy and reduce noise. Peak width was manually adjusted to match the isotopically resolved data. Spectra were sampled at every 0.1 Da. Charge state distributions were automatically smoothed. Deconvoluted spectra were quantified by calculating the weight average mass across a mass region spanning all observed modifications for a given intact protein.

2.5.5 Peptide mass spectrometry and modification analysis

Prior to digestion, samples were diluted 5-fold with 100 mM Tris pH 8.0. Trypsin (Promega) was added at a protease:protein ratio of 1:100 and samples were incubated overnight at 37 °C. Digestion was quenched with addition of formic acid to a final concentration of 1%. Following digestion, peptides were loaded at a flow rate of 50 uL/min in 99% Buffer A (0.1% Formic Acid) 1% buffer B (100% Acetonitrile, 0.1% Formic Acid) onto a Pepmap C18 column (Thermo Fisher 164711) stored in a column oven maintained at 50 °C. Residual buffer and salts were removed by maintaining flow for 4 minutes with a valve diverting to waste. Peptides were then eluted with a gradient of 1-40% buffer B at a flow rate of 50 µL/min over 15 minutes, and then of 40-90% buffer B over 30 seconds. The column was then subjected to a sawtooth wash and equilibrated at 1% buffer B prior to the next injection. Peptides were eluted directly into a Q Exactive Orbitrap Mass Spectrometer operating in positive mode (HESI source settings: Spray voltage 3.5 kV, Capillary temp. 150 C, S-lens RF level 70, Aux gas heater temp. 150 °C — MS Settings: resolution 70000, AGC target 3e6, maximum IT 50 ms, scan range 150-2000 m/z — dd-MS^2 settings: resolution 17500, AGC target 1e5, maximum IT 50 ms, loop count 10, isolation window 2.6 m/z, NCE 28, charge state 1 and ≥7 excluded, dynamic exclusion of 2 seconds). LC and MS methods were run using Xcalibur 4.1 (Thermo Scientific)

Modified and unmodified peptide identification and quantification were performed using the Byos Oxidative Footprinting platform (Protein Metrics Incorporated) incorporating both the Byonic search engine and Byologic quantification program. All peptide experiments with fragmentation data were searched for tryptic peptides originating from the search library. The search library consisted of a fasta file containing sequences for all proteins present in solution. Half tryptic peptides and peptides with missed tryptic cleavages were present, but with lower intensity, and were excluded from further analysis. For a given tryptic peptide, all detected modifications were quantified if they were observed in any sample. For each modification type (i.e. +16 Da, +32 Da, +48 Da, etc.) was separately quantified by summing the area of all chromatographic peaks with that mass, regardless of whether an MSMS identification was made at that retention time. Chromatographic peaks with isotopic distribution that did not match the unmodified peak, or for which MSMS had identified as a different peptide, were not quantified. The final percent modified was the sum of all modified peak areas divided by the sum of all modified and unmodified peak areas.

2.5.6 Circular dichroism

Circular dichroism experiments were performed with an Aviv 410 spectrometer. Each sample contained 40 µg/mL protein (either ecRNH*, T4L*, or an equimolar mixture of both) with varying concentrations of urea. All CD melts were performed in 20 mM NaOAc 50 mM KCl pH 5.5. The signal at 222 nm was measured in a 1-cm path length Starna Cells cuvette at 25°C with stirring, averaged over one minute.

2.5.7 Extraction of thermodynamic parameters from denaturation melts

Denaturant melts, for both circular dichroism and HRF-MS, were analyzed using the linear extrapolation model, resulting in both a Cm and m-value describing the unfolding transition, along with parameters describing the native and denatured baselines. For data with linear baselines, the following equation was used:

$$(Y_N + B_N[Urea]) + \frac{\left(Y_D + B_D[Urea]\right)e^{m^*([Urea]-C_M)/(RT)}}{1+e^{m^*([Urea]-C_M)/(RT)}}$$

Where m and C_m are the m-value and concentration midpoint, Y_D and Y_N represent the denatured and native baseline Y-intercepts, B_D and B_N represent the denatured and native baseline slopes, and where R is the gas constant and T is the temperature in Kelvin. R and T were fixed at 1.987E-3 kcal/mol/K and 293 K respectively.

For peptide HRF-MS melts with apparent exponential native baselines, the following equation was used:

$$(A_N e^{-k_N[Urea]} + B_N) + \frac{\left(Y_D + B_D[Urea]\right) e^{-m^*(x-C_M)/(RT)}}{1+e^{-m^*(x-C_M)/(RT)}}$$

With the same definitions as above and where A_N, B_N, and k_n define the native baseline.

Parameters were determined by non-linear least squares fitting of the data to the appropriate model in Igor Pro.

2.6 References

1. Dill, K. A. *et al.* Principles of protein folding--a perspective from simple exact models. *Protein Sci.* **4**, 561–602 (1995).

2. Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **338**, 1042–1046 (2012).

3. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.* **37**, 289–316 (2008).

4. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

5. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

6. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).

7. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

8. Street, T. O., Courtemanche, N. & Barrick, D. Protein Folding and Stability Using Denaturants. in vol. 84 295–325 (Elsevier, 2008).

9. Thompson, S., Zhang, Y., Ingle, C., Reynolds, K. A. & Kortemme, T. Altered expression of a quality control protease in E. coli reshapes the in vivo mutational landscape of a model enzyme. *Elife* **9**, (2020).

10. Stein, A., Fowler, D. M., Hartmann-Petersen, R. & Lindorff-Larsen, K. Biophysical and Mechanistic Models for Disease-Causing Protein Variants. *Trends Biochem. Sci.* (2019) doi:10.1016/j.tibs.2019.01.003.

11. Faure, A. J. *et al.* Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022).

12. Park, C. & Marqusee, S. Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. *Nat. Methods* **2**, 207–212 (2005).

13. Park, C. & Marqusee, S. Quantitative determination of protein stability and ligand binding by pulse proteolysis. *Curr. Protoc. Protein Sci.* **Chapter 20**, Unit 20.11–20.11.14 (2006).

14. Ghaemmaghami, S., Fitzgerald, M. C. & Oas, T. G. A quantitative, high-throughput screen for protein stability. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8296–8301 (2000).

15. Dai, S. Y. & Fitzgerald, M. C. Accuracy of SUPREX (stability of unpurified proteins from rates of H/D exchange) and MALDI mass spectrometry-derived protein unfolding free energies determined under non-EX2 exchange conditions. *J. Am. Soc. Mass Spectrom.* **17**,

1535–1542 (2006).

16. West, G. M., Tang, L. & Fitzgerald, M. C. Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy. *Anal. Chem.* **80**, 4175–4185 (2008).

17. Walker, E. J., Bettinger, J. Q., Welle, K. A., Hryhorenko, J. R. & Ghaemmaghami, S. Global analysis of methionine oxidation provides a census of folding stabilities for the human proteome. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6081–6090 (2019).

18. Maxwell, K. L. *et al.* Protein folding: defining a standard set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**, 602–616 (2005).

19. Gupta, S., Feng, J., Chance, M. & Ralston, C. Recent Advances and Applications in Synchrotron X-Ray Protein Footprinting for Protein Structure and Dynamics Elucidation. *Protein Pept. Lett.* **23**, 309–322 (2016).

20. Wang, L. & Chance, M. R. Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Anal. Chem.* **83**, 7234–7241 (2011).

21. Dabora, J. M. & Marqusee, S. Equilibrium unfolding of Escherichia coliribonuclease H: Characterization of a partially folded state. *Protein Sci.* **3**, 1401–1408 (1994).

22. Chamberlain, A. K., Handel, T. M. & Marqusee, S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Biol.* **3**, 782–787 (1996).

23. Hollien, J. & Marqusee, S. A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* **38**, 3831–3836 (1999).

24. Samelson, A. J., Jensen, M. K., Soto, R. A., Cate, J. H. D. & Marqusee, S. Quantitative determination of ribosome nascent chain stability. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13402–13407 (2016).

25. Raschke, T. M., Kho, J. & Marqusee, S. Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat. Struct. Biol.* **6**, 825–831 (1999).

26. Myers, J. K., Nick Pace, C. & Martin Scholtz, J. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).

27. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M. & Fersht, A. R. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **32**, 11259–11269 (1993).

28. Hu, W. *et al.* Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7684–7689 (2013).

29. Connell, K. B., Miller, E. J. & Marqusee, S. The Folding Trajectory of RNase H Is

Dominated by Its Topology and Not Local Stability: A Protein Engineering Study of Variants that Fold via Two-State and Three-State Mechanisms. *J. Mol. Biol.* **391**, 450–460 (2009).

30. Mateus, A., Määttä, T. A. & Savitski, M. M. Thermal proteome profiling: unbiased assessment of protein state through heat-induced stability changes. *Proteome Sci.* **15**, 13 (2016).

31. Mateus, A. *et al.* Thermal proteome profiling for interrogating protein interactions. *Mol. Syst. Biol.* **16**, e9232 (2020).

32. Franken, H. *et al.* Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* **10**, 1567–1593 (2015).

33. Leuenberger, P. *et al.* Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, (2017).

34. Schopper, S. *et al.* Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat. Protoc.* **12**, 2391–2410 (2017).

35. Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044 (2014).

36. Piazza, I. *et al.* A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **172**, 358–372.e23 (2018).

37. To, P., Whitehead, B., Tarbox, H. E. & Fried, S. D. Nonrefoldability is Pervasive Across the E. coli Proteome. *J. Am. Chem. Soc.* **143**, 11435–11448 (2021).

38. Sharp, J. S. *et al.* Flash Oxidation (FOX) System: A Novel Laser-Free Fast Photochemical Oxidation Protein Footprinting Platform. *J. Am. Soc. Mass Spectrom.* (2021) doi:10.1021/jasms.0c00471.

39. Gupta, S., Celestre, R., Feng, J. & Ralston, C. Advancements and Application of Microsecond Synchrotron X-ray Footprinting at the Advanced Light Source. *Synchrotron Radiat. News* **29**, 39–44 (2016).

40. Gupta, S., Celestre, R., Petzold, C. J., Chance, M. R. & Ralston, C. Development of a microsecond X-ray protein footprinting facility at the Advanced Light Source. *J. Synchrotron Radiat.* **21**, 690–699 (2014).

41. Kostelic, M. M. & Marty, M. T. Deconvolving Native and Intact Protein Mass Spectra with UniDec. in *Proteoform Identification: Methods and Protocols* (eds. Sun, L. & Liu, X.) 159–180 (Springer US, 2022).

42. Marty, M. T. *et al.* Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

Chapter 3

The SARS-CoV-2 spike reversibly samples an open-trimer conformation exposing novel epitopes

---

This chapter is adapted from the paper:
Costello SM, Shoemaker SR, Hobbs HT, Nguyen AW, Hsieh C-L, Maynard JA, McLellan JS, Pak JE, Marqusee S. 2022. The SARS-CoV-2 spike reversibly samples an open-trimer conformation exposing novel epitopes. *Nature Structural & Molecular Biology.* 29, 229–238.

I am the co-first author of this work. My contributions to this work include the conceptualization of the project, performance of experimental work, data analysis and interpretation, and writing of the manuscript.

3.1 Abstract

Current COVID-19 vaccines and many clinical diagnostics are based on the structure and function of the SARS-CoV-2 spike ectodomain. Using hydrogen-deuterium exchange mass spectrometry, we have uncovered that, in addition to the prefusion structure determined by cryo-EM, this protein adopts an alternative conformation that interconverts slowly with the canonical prefusion structure. This new conformation—an open trimer—contains easily accessible RBDs. It exposes the conserved trimer interface buried in the prefusion conformation, thus exposing potential epitopes for pan-coronavirus antibody and ligand recognition. The population of this state and kinetics of interconversion are modulated by temperature, receptor binding, antibody binding, and sequence variants observed in the natural population. Knowledge of the structure and populations of this conformation will help improve existing diagnostics, therapeutics, and vaccines.

3.2 Introduction

The spike protein from SARS-CoV-2 (also referred to as the S-protein) is the primary target for current vaccines against COVID-19 and the focus of many therapeutic efforts (*1–4*). This large heavily glycosylated trimeric protein is responsible for cell entry via recognition of the host receptor angiotensin-converting enzyme 2 (ACE2) and membrane fusion (*5–7*). It is also the principal antigenic determinant of neutralizing antibodies (*8*). Shortly after release of the viral genome sequence, a version of the spike ectodomain (termed S-2P) was designed to stabilize the prefusion conformation, and the structure was determined by cryo-electron microscopy (cryo-EM) (*9*, *10*).

This S-2P ectodomain comprises the first ~1200 residues of the spike protein (Fig. 3-1A) with two proline substitutions in the S2 domain designed to stabilize the prefusion conformation, mutations that abolish the furin-cleavage site, and the addition of a C-terminal trimerization

motif (*9*). This mimic, its structure, and others that followed, have been widely used for vaccine development and interpretation of many structure/function and epidemiological studies. To date there are more than 250 structures of SARS-CoV-2 spike ectodomains in the Protein Data Bank (*11*).

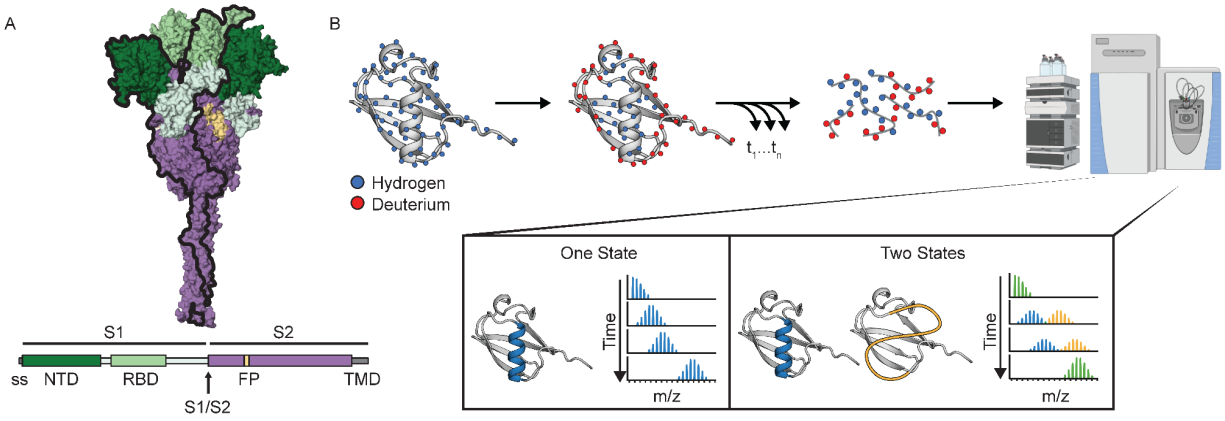


Figure 3-1. SARS-CoV-2 spike ectodomain and Hydrogen-Deuterium Exchange monitored by mass spectrometry (HDX-MS) experimental overview. (A) Schematic of the prefusion-stabilized SARS-CoV-2 spike protein and a model of the trimeric prefusion conformation (24). **(B)** Schematic of HDX-MS experiment and the resulting mass distributions for a peptide that exists in either one (left) or two (right) separable conformations. In order for the two conformations to result in a bimodal mass distribution, they must not interconvert during the timescale of the HDX experiment (hours). Rapid interconversion would result in a single mass distribution with the ensemble averaged mass profile.

These structural studies together with other functional studies demonstrate that, like all class 1 viral fusion proteins, the spike protein is dynamic, sampling several different conformations during its functional lifecycle (*12*, *13*). The three individual receptor-binding domains (RBDs) sample an 'up' state and a 'down' state; the up state exposes the ACE2-binding motif and therefore is required for infectivity (*7*, *10*, *14*, *15*). After receptor binding and cleavage between the S1 and S2 domains, the protein undergoes a major refolding event to allow fusion and adopts the stable post-fusion conformation (*6*, *7*, *16–18*).

Despite the wealth of structural information, there are very few experimental studies on the dynamics within the prefusion state. The noted RBD up/down conformational transition has been monitored on the membrane via single molecule FRET and occurs on the order of seconds (*19*). Large computational resources have been devoted to molecular simulations of the spike protein revealing a dynamic pre-fusion state with a range of accessible conformations including the potential of a further opening of the RBD and N-terminal domain (NTD) away from the trimer interface (*20*, *21*). Experimentally, the conformational landscape of spike has not been well interrogated and the effects of perturbations, such as ligand binding (both receptor and antibodies) or amino acid substitutions found in emerging variants of concern are unknown.

For these reasons, we turned to hydrogen deuterium exchange monitored by mass spectrometry (HDX-MS) to probe the conformational landscape of the soluble spike prefusion ectodomain as well as the effects of ligand binding and sequence variation. We uncovered a stable alternative conformation that interconverts slowly with the canonical prefusion structure. This conformation is an open trimer, with easily accessible RBDs. It exposes the S2 trimer interface, providing new epitopes in a highly conserved region of the protein.

3.3 Results

3.3.1 Continuous exchange HDX-MS on Spike 2P (S-2P)

HDX-MS offers an ideal complement to the ever-growing number of structural studies on the SARS-CoV-2 spike protein, providing information on its conformational ensemble and dynamics. HDX-MS monitors the time course of exchange of amide hydrogens on the peptide backbone with the hydrogens in the solvent (see Fig. 3-1B for description). An individual amide's ability to undergo exchange is directly related to its structure and stability (*22*, *23*).

We first followed the continuous exchange time course of hydrogen exchange at 25 °C on the entire S-2P ectodomain, over a period of 15 seconds to 4 hours (see Materials and Methods). Using a combination of porcine pepsin and aspergillopepsin digestion, we obtained 85% peptide coverage allowing us to interrogate the dynamics of the entire protein (800 peptides, which include 9 of the 22 glycosylation sites, average redundancy of 8.6) (Fig. 3-2A). Notably, we have coverage in areas not resolved in the cryo-EM structure, including loops in the N-terminal domain (NTD) and RBD that have been found to be recognized by antibodies, loops in the S2 region that include the protease cleavage sites, and C-terminal residues after residue 1145 which includes the second heptad repeat (HR2). Based on control experiments using deuterated protein, our HDX protocol results in an average back exchange of 22% (Fig. 3-2B).

The vast majority of peptides show a classic single isotopic envelope whose centroid increases in mass as deuterons are added over time (Fig. 3-1B). A small minority of the peptides, however, show bimodal behavior—with two isotopic envelopes both increasing in mass over time: one less-exchanged distribution and a second more-exchanged distribution (these peptides are described in detail below). The HDX profile of all the peptides, with the exception of the more-exchanged distributions in the bimodal peptides, is consistent with the known prefusion conformation (Fig. 3-3, Fig. 3-4): secondary structure and buried elements within the trimer exchange slower than exposed loops. We also observe protection for residues 1140–1197, which includes HR2, a region not defined in single-particle cryo-EM structures, supporting the predicted helical structure of this region (*24*) and the relative rigidity of the stalk observed by cryo-electron tomography (cryo-ET) (*25*).

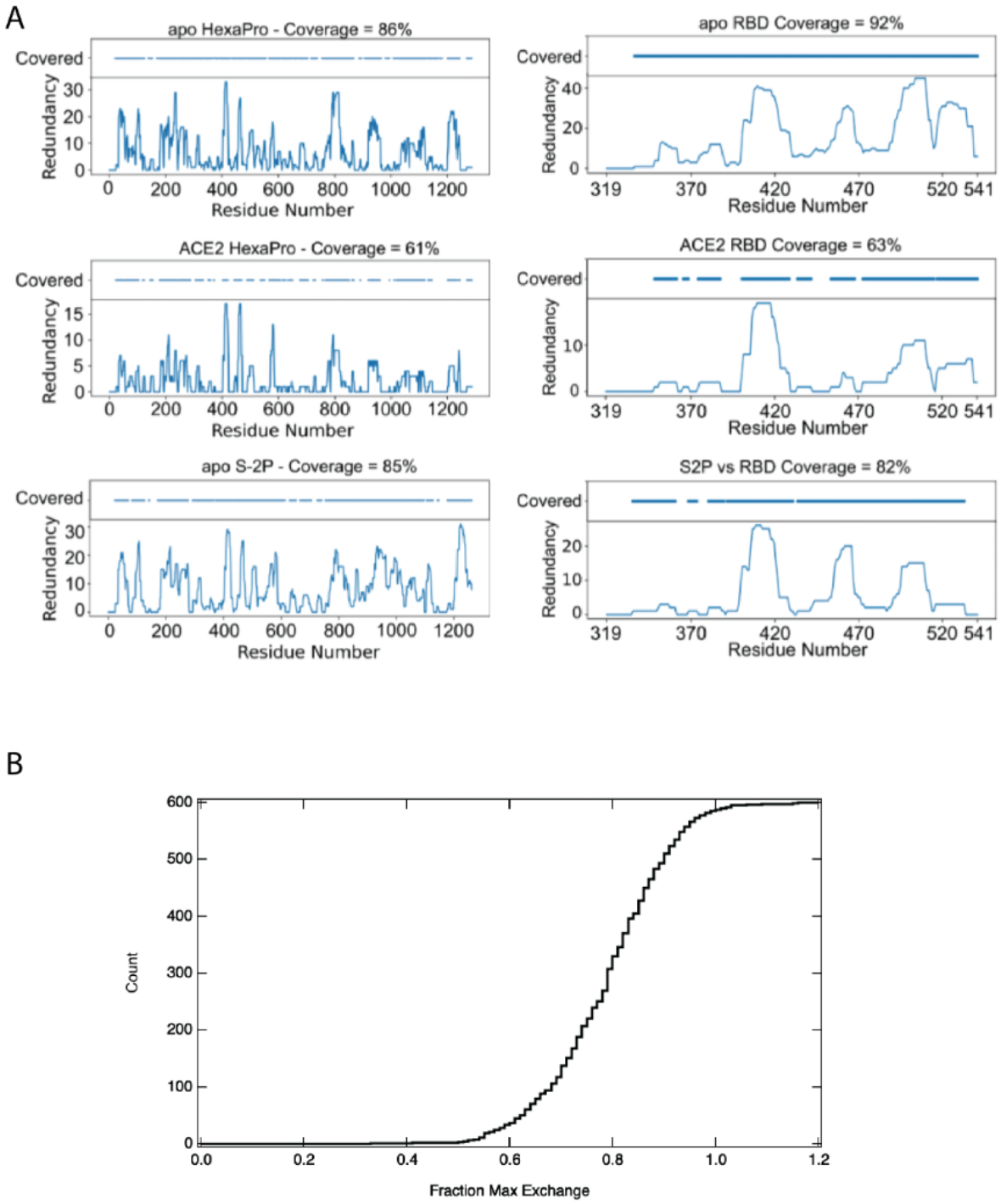


Figure 3-2. Coverage, redundancy, and back exchange results from HDX-MS experiments
(A) Peptide coverage and redundancy at each residue for all HDX-MS experiments. **(B)** Back-Exchange Control: Cumulative histogram of the fractional deuterium maintained during workup of a fully deuterated sample. Fraction max exchange is corrected for the 90% D_2O experimental conditions.

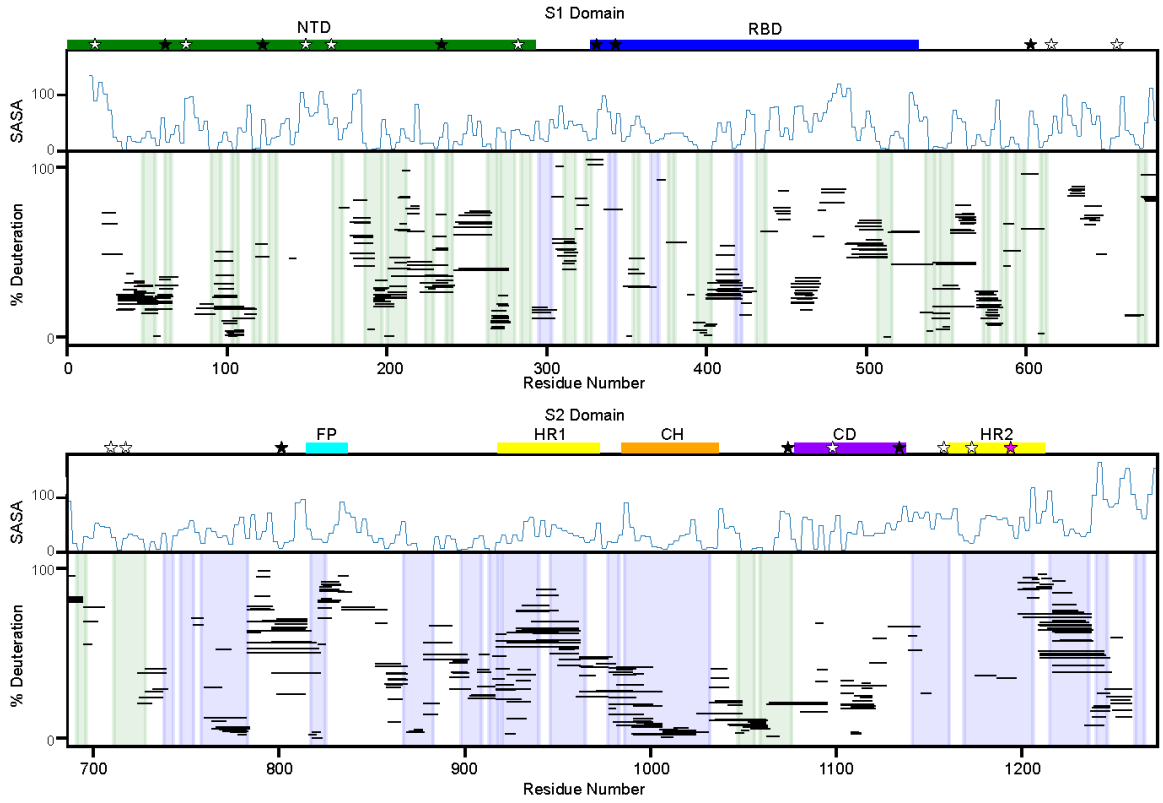
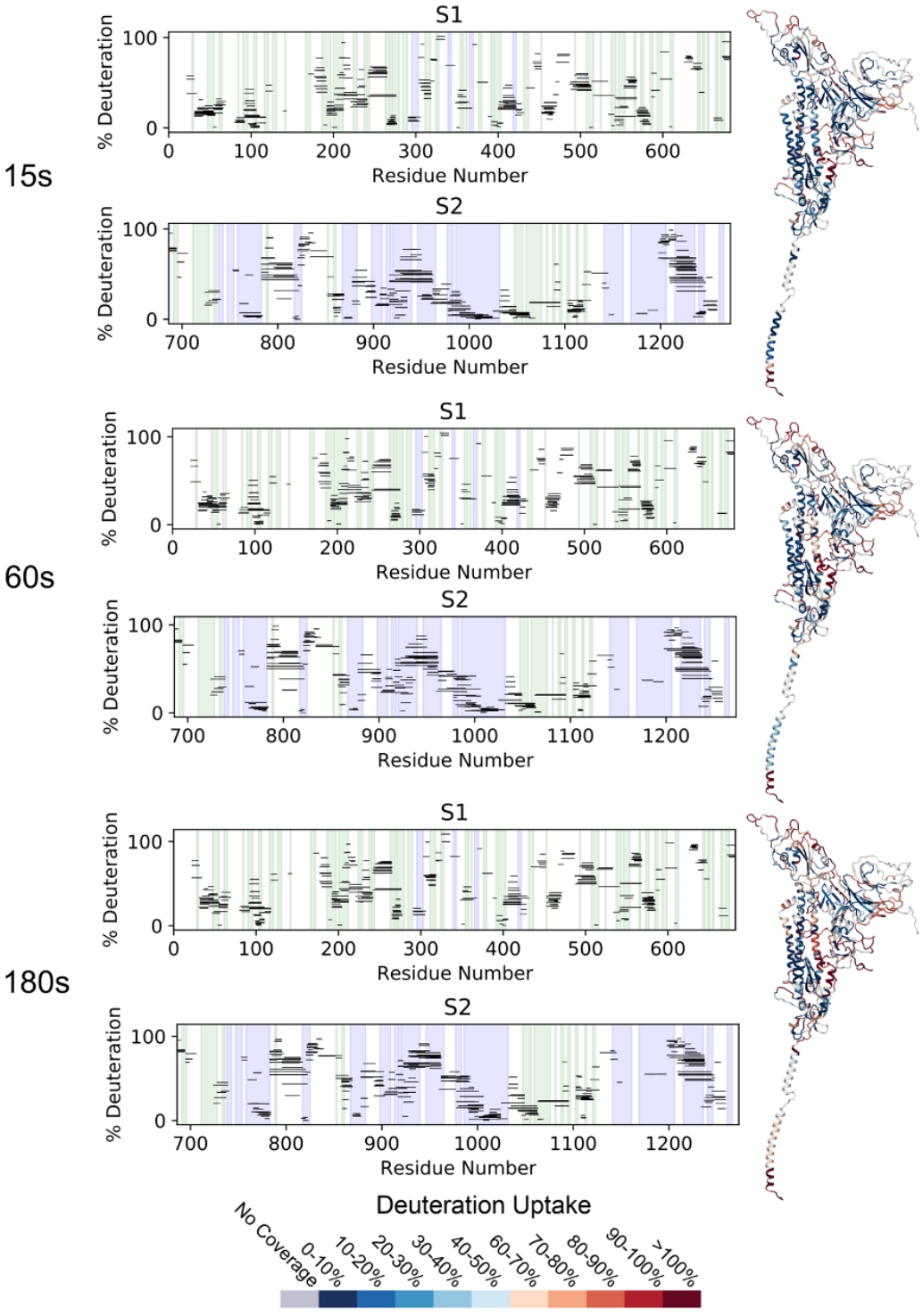
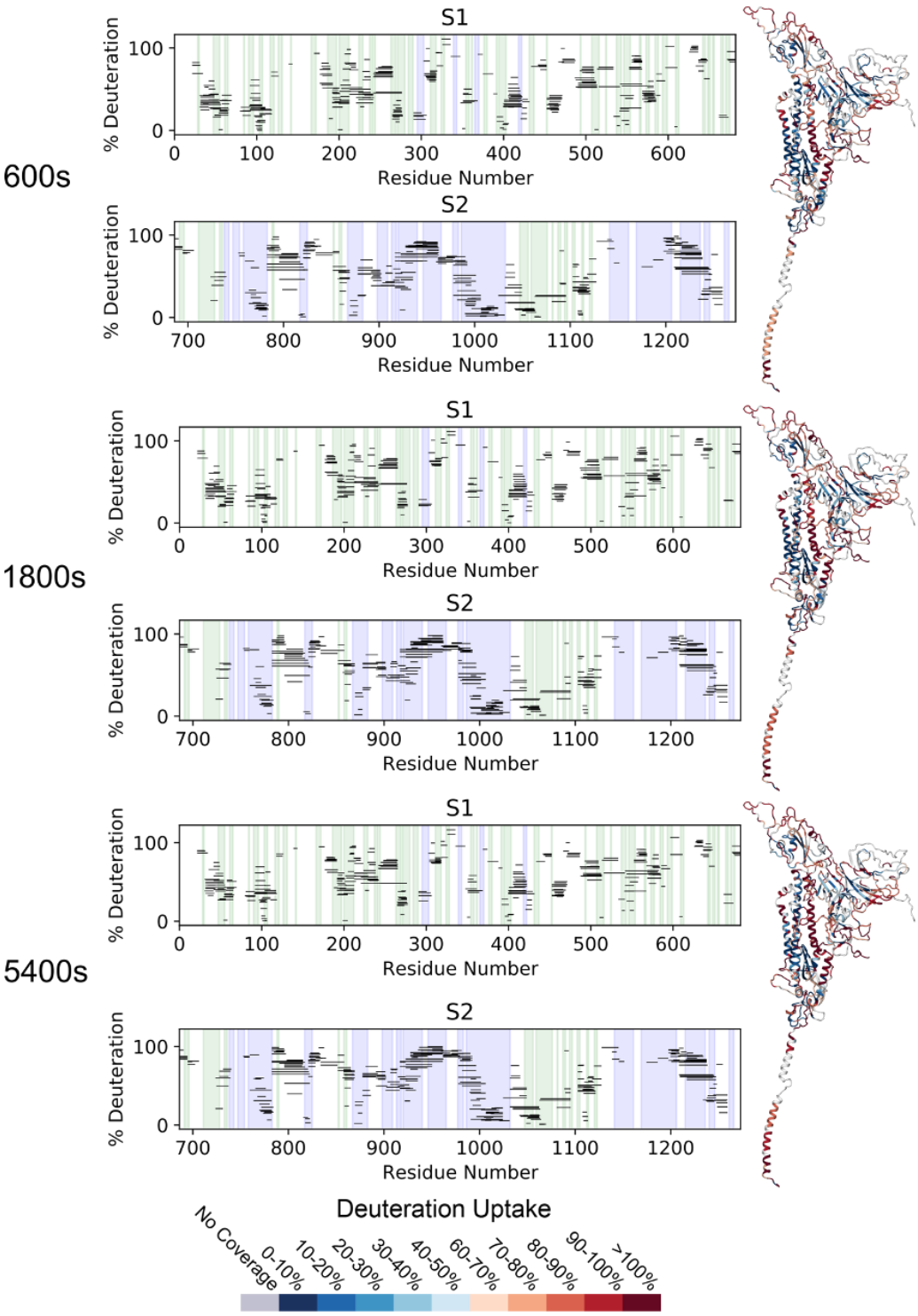


Figure 3-3. Peptide-level HDX is consistent with the known prefusion conformation. Percent deuteration after one minute of deuterium labeling for every peptide in the S-2P continuous exchange dataset for the S1 (top) and S2 (bottom) domains. Each line represents an individual peptide spanning the residues indicated on the x-axis, with percent deuteration after one minute of exchange indicated on the y-axis (for bimodal peptides, only the less-exchanged centroid shown). Secondary structures in the prefusion conformation are shaded in blue (alpha helices) and green (beta strands). A measure of solvent accessibility is shown above in $Å^2$ (calculated as a three-residue sliding average using the default get_area function in pymol) using a model of the full-length prefusion trimer with all three RBDs in a down position (24). These data are consistent with the SARS-CoV-2 spike trimer secondary structures, notably regions buried in the trimer interface, such as the central helix, show increased protection relative to more exposed regions lacking secondary structure. Important sequence features are indicated above the plot including the N-terminal domain (NTD - green), receptor binding domain (RBD - blue), fusion peptide (FP - cyan), heptad repeat 1 (HR1 - yellow), central helix (CH - orange), core domain (CD - purple) and heptad repeat 2 (HR2 - yellow). Locations of glycans are noted with stars with three categories - glycans detected in at least one peptide of our data set (black, 9/22), glycans known to be on the spike protein where we lack coverage (white, 12/22), and glycans known to be on the spike protein but for which non-glycosylated peptides are observed (pink, 1/22).


15s
S1
% Deuteration
100
0
0
100
200
300
400
500
600
Residue Number
S2
% Deuteration
100
0
700
800
900
1000
1100
1200
Residue Number
60s
S1
% Deuteration
100
0
0
100
200
300
400
500
600
Residue Number
S2
% Deuteration
100
0
700
800
900
1000
1100
1200
Residue Number
180s
S1
% Deuteration
100
0
0
100
200
300
400
500
600
Residue Number
S2
% Deuteration
100
0
700
800
900
1000
1100
1200
Residue Number
Deuteration Uptake
No Coverage
0-10%
10-20%
20-30%
30-40%
40-50%
60-70%
70-80%
80-90%
90-100%
>100%


600s
1800s
5400s
S1
S2
% Deuteration
Residue Number
Deuteration Uptake
No Coverage
0-10%
10-20%
20-30%
30-40%
40-50%
60-70%
70-80%
80-90%
90-100%
>100%

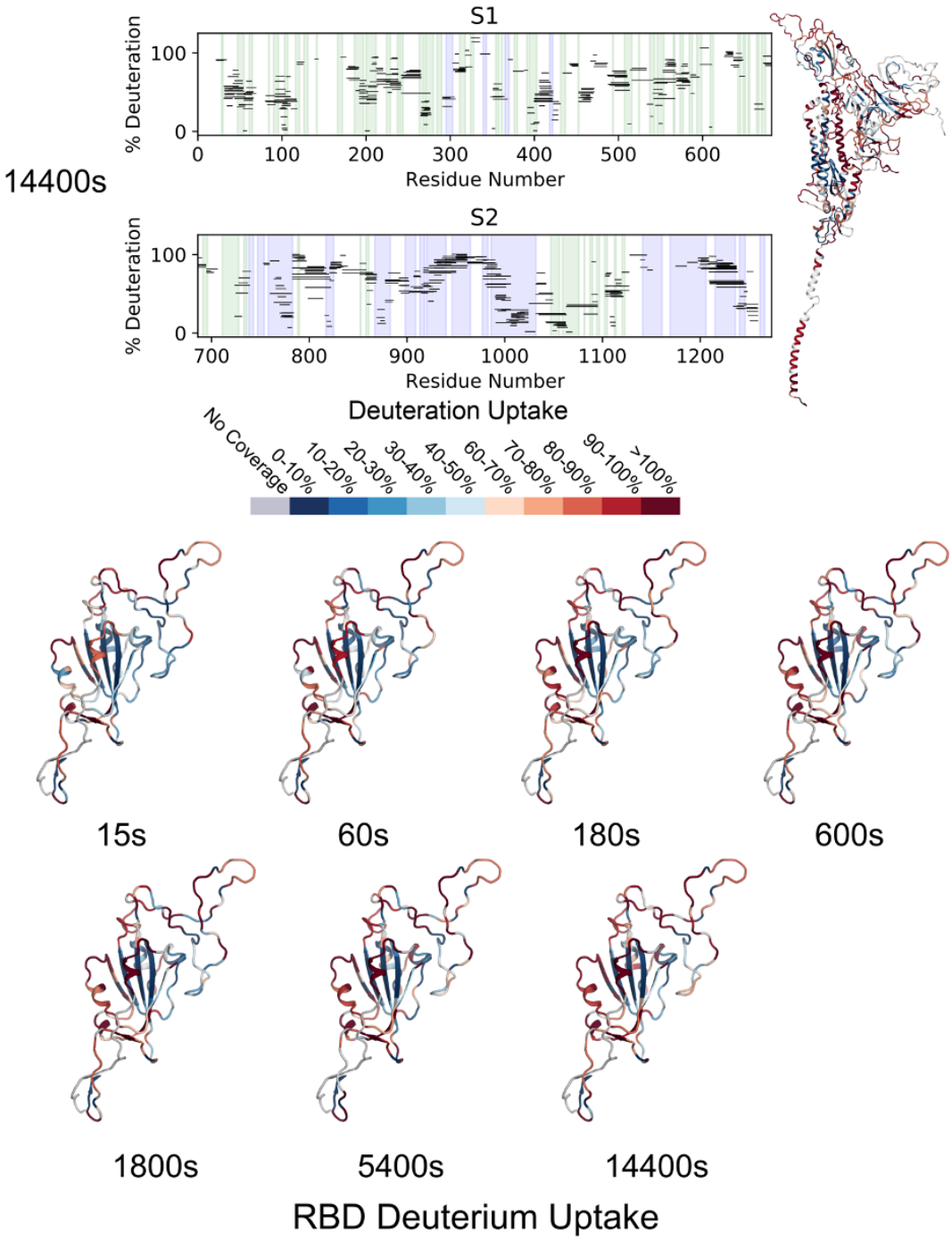


Figure 3-4. Spike HDX-MS results as a function of time. (Top) Deuterium uptake for each S-2P experimental time point (left) shown as a function of sequence position for every peptide analyzed (center) and with per-residue deuteration uptake (scale shown at bottom) mapped to the structure a single protomer of a full length prefusion spike trimer model from (24) (right). Secondary structures in the prefusion structure are shaded in blue (alpha helices) and green (beta strands). (Bottom) Per-residue deuterium uptake for each Apo-RBD experimental time point mapped to the structure of the RBD (single RBD from a full-length spike trimer model from (24)). Per-residue deuteration calculated from all peptide data by HDExaminer 3.

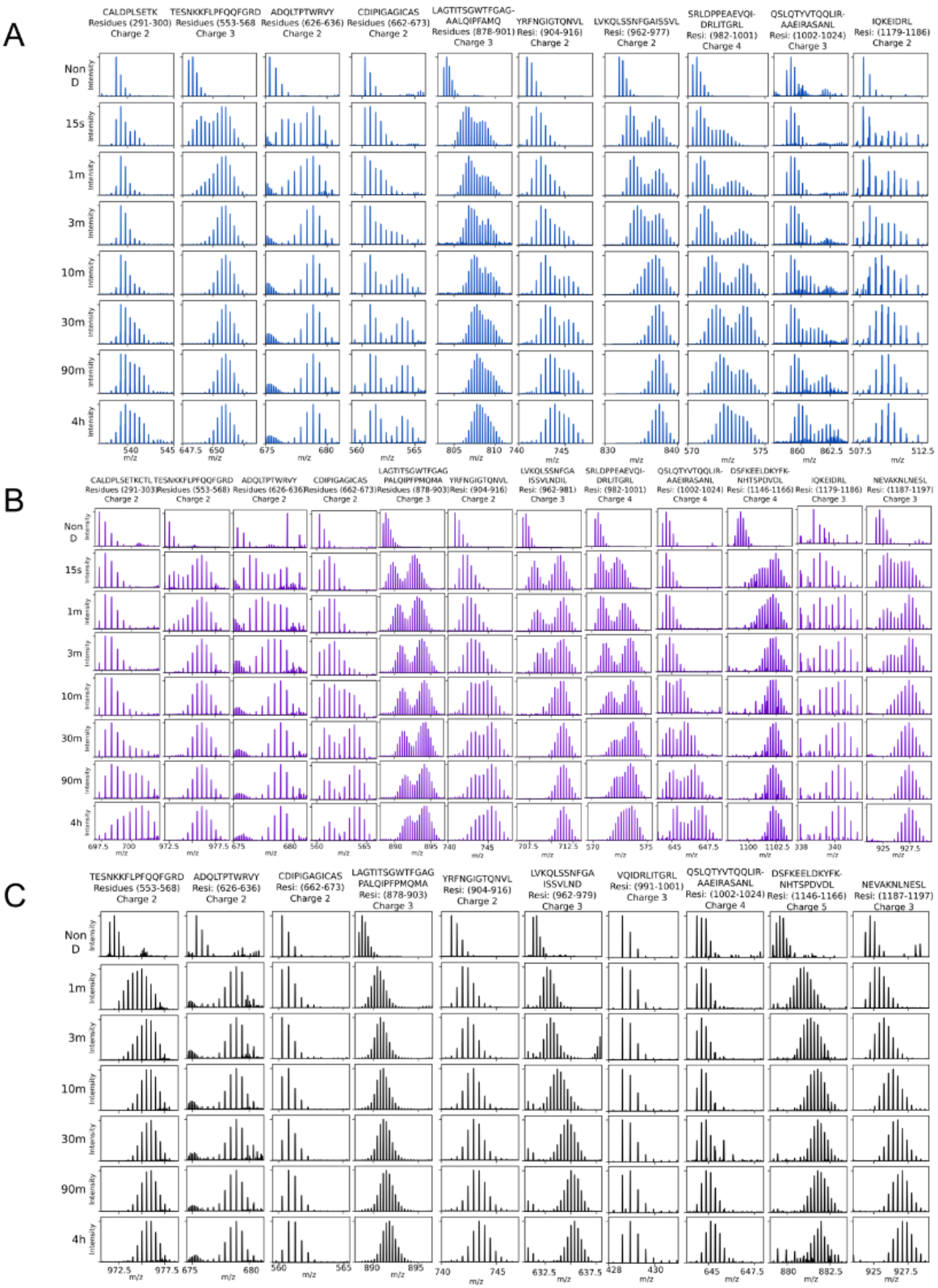


Figure 5. Bimodal peptide spectra observed in continuous labeling HDX-MS experiments Observed spectra of a representative set of peptides with observed bimodal behavior for **(A)** S-2P and **(B)** HexaPro **(C)** Disulfide-locked HexaPro

3.3.2 Identification of an alternative conformation

Bimodal mass envelopes can indicate the presence of two different conformations that interconvert slowly on the timescale of our hydrogen exchange experiment: one where the amides are more accessible to exchange compared to the other. However, it can also be a result of the kinetics of the hydrogen exchange process itself, so-called EX1 exchange (when the rates of hydrogen bond closing are much slower than the intrinsic chemistry of the exchange process). In this rare scenario, the heavier mass distribution will increase in intensity at the expense of the lighter one over the observed time period. This is not what we observe for the spike protein: the bimodal mass distributions retain their relative intensities, increasing in average mass over time (Fig. 3-5). The observed ratio is the same for every bimodal peptide under any given condition. The fact that every peptide shows the same ratio indicates that these bimodal peptides reflect two conformations; they report on the regions of the protein that show differences in hydrogen exchange in each conformation.

The bimodal peptides we observe are predominantly in the most conserved region of the protein—the S2 region (*26*) (Fig. 3-6A). When mapped onto the canonical prefusion conformation, many come from the helices at the trimer interface (residues 962–1024, 1146–1166, 1187–1196), indicating a second conformation with less stable hydrogen bonding for these helices, consistent with a loss of interprotomer contacts and increased solvent accessibility. We also observe bimodal peptides in other areas of the inter-protomer interface, such as residues 870–916 in S2 and residues 553–574 and 662–673 in S1, again suggesting a change in trimer contacts. Finally, we see bimodal peptides in two regions that do not form interprotomer contacts (residues 291–305, 626–636); instead, these residues form the interface between the NTD and second S1 subdomain (SD2), suggesting that this subdomain interface is also structurally different in this second conformation. All the other peptides (the majority) fit to a classic unimodal distribution in the mass spectrum. The fact that the majority of the peptides are unimodal and that they behave the same in both conformations, indicate that they exist in similar structures in each conformation and suggest that the individual domains have the same structure in each conformation. Previous HDX studies involving the spike protein did not note this behavior, which can be attributed to differences in the experimental conditions and protocols. (*27*, *28*) (see Supplementary Text).

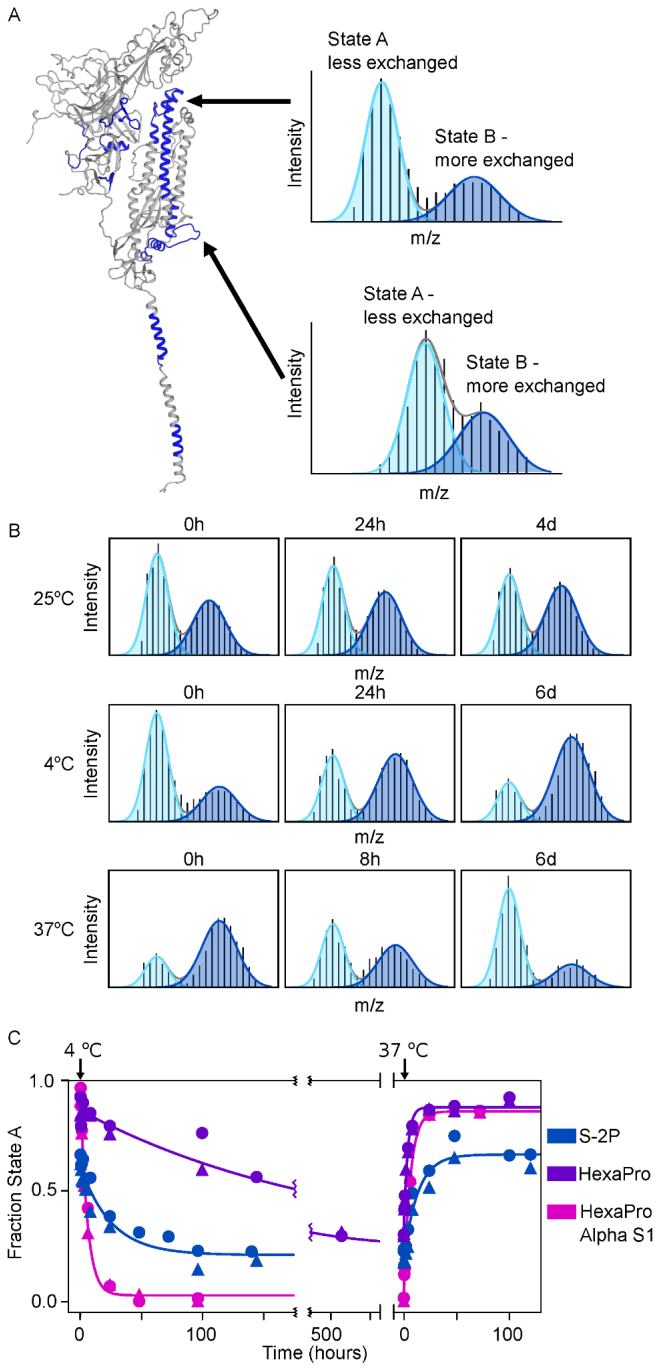


Figure 3-6. The Spike ectodomain reversibly samples two conformations. **(A)** Left: SARS-CoV-2 spike monomer with all regions that have peptides showing bimodal mass distributions colored in blue. Right: Example mass spectra from two peptides after one minute of deuteration (top: residues 982–1001, bottom: residues 878–903) with overlaid fitted gaussian distributions that describe each protein conformation in blue (light blue: less exchanged A state, dark blue: the more exchanged B state). **(B)** Conformational preference for the S-2P spike construct at 25 °C, 4 °C and 37 °C monitored by pulsed-labeling. At 25 °C S-2P converts from primarily state A to ~50:50 A:B after 4 days. At 4 °C, S-2P prefers state B while at 37 °C, S-2P prefers state A. **(C)** The kinetics of interconversion between the A and B states for different of spike variants. Starting from an initial prefusion conformation (state A, 37 °C), samples were rapidly transferred to 4 °C and assayed for conversion to state B over time using pulsed-labeling HDX-MS. To estimate fraction state A, peptides from two different regions (residues 982-1001 (circles) and residues 878-903 (triangles)) were fit to two gaussians. Data from both regions were used to determine the rate of interconversion.

3.3.3 Interconversion between the two conformations

These data suggest a model where the spike protein populates two conformations within the prefusion state—the classical prefusion structure seen in cryo-EM (herein referred to as state A)—and one where each domain has a similar protomer topology and a more flexible or exposed open-trimer interface (herein referred to as state B). The above data do not, however, provide evidence that these states interconvert; any potential interconversion must be slower than the four-hour hydrogen-exchange experiment. Since the transition of the RBD between the up and down conformation occurs on the order of seconds, this conformational heterogeneity is not the source of the bimodal distributions and the observed hydrogen exchange reports on the weighted average of the two RBD conformations. There are several irreversible situations that could account for conformational heterogeneity such as differences in glycosylation, proteolytic degradation, irreversible misfolding, or aggregation. To rule these out, we tested whether the two conformations interconvert reversibly. We used bimodal peptides to quantify the population of each conformation under differing conditions (such as temperature, time, ligand, etc.). Under each condition, we carried out a one-minute pulse of hydrogen exchange (25 °C) and integrated the area under the two mass envelopes for a single bimodal peptide to ascertain the fraction of each conformation under that condition or moment in time (see Materials and Methods). We monitored two bimodal peptides, one from region 878-903 and one from region 978-1001. Using these two peptides, we quantified the population of each conformation under different conditions (*29*), such as temperature, time, and ligand. We chose these bimodal peptides because of their high signal to noise ratio and because they report on two distinct regions, thus providing information from both the top and bottom of the S2 interface. We selected a pulse length of one-minute as it provides clearly separable bimodal distributions for these two peptides. For every condition tested, irrespective of the A:B ratio, both peptides report the same fractional population for the two conformations, indicating that all these data can be best described as a variable mixture of just two conformations: the canonical prefusion conformation and an unexpected alternative conformation.

Long-term incubation (four days, 25 °C, pH 7.4) demonstrated a slow shift in population from a majority in the canonical prefusion state (state A) to a majority in the alternative conformation (state B) (Fig. 3-6B, Fig. 3-7A). Thus, the prefusion state can transform into the alternative state and the bimodal behavior cannot be due to sample heterogeneity such as differential glycosylation. This observed A → B conversion, however, does not rule out an irreversible process such as degradation or misfolding.

Postulating that the bimodal peaks represent a reversible structural transition, we used temperature to perturb the system and investigate the ability to interconvert. Indeed, the conformations do interconvert reversibly, with a preference for B at 4 °C and A at 37 °C (Fig. 3-6B). The observed kinetics of interconversion are extremely slow: A → B $t_{1/2}$ of ~17 hours at 4 °C and, when that same sample is moved to ~37 °C, B → A $t_{1/2}$ of ~ 9 hours (Fig. 3-6C, Table 3-1). Notably the final, and presumably equilibrium, distribution at either temperature shows an observable population of both states (>5%), indicating an energy difference of less than 2 kcal/mol between the two conformations under both conditions.

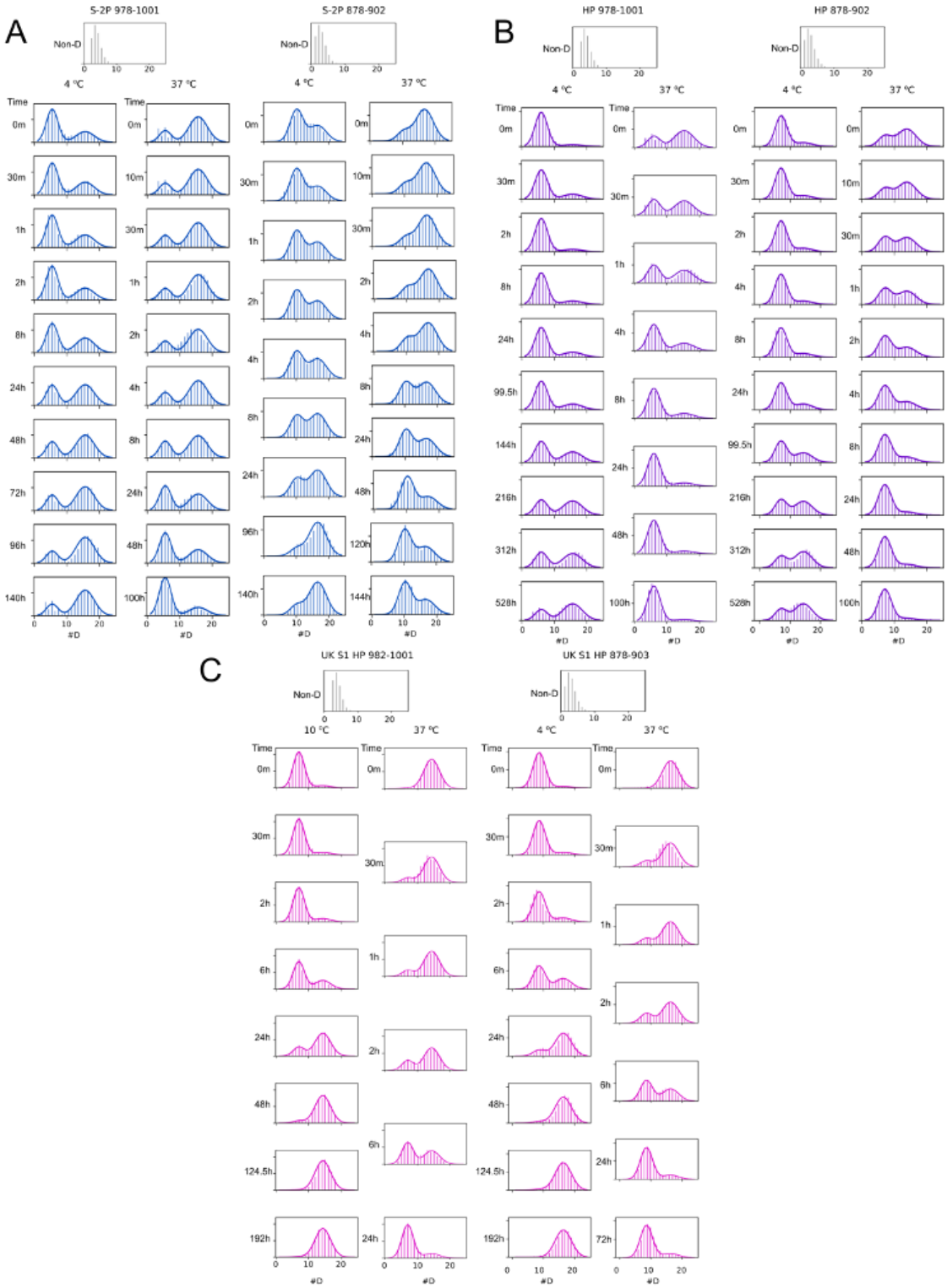


Figure 3-7. Example bimodal peptide spectra observed in pulsed-labeling HDX-MS experiments One example time course for a peptide from each of the two regions used to quantify states A and B in (A) S-2P, (B) HexaPro, and (C) alpha S1 HexaPro. The bimodals in the time course were globally fit to a sum of two gaussians to determine the distributions for state A and state B, the fits were then used to quantify the relative populations of state A and state B (see methods), the resulting gaussian fits are overlaid. Undeuterated spectra are shown at the top in gray.

Table 3-1. Rates of interconversion between the A (prefusion) and B (open trimer) conformations of spike ectodomains. $k_{observed}$ is the observed rate of change in the population of the A state after a temperature jump. This relaxation rate is the sum of the forward and reverse rates, which is dominated by the major conformational change (A→B at 4 °C, 10 °C and B→A at 37°C). $t_{1/2}$ is the half time for that same rate, $\ln 2/k_{obs}$).

Temperature	Protein	$k_{observed}$ (hr^{-1})	$t_{1/2}$ (hours)
37 °C→4 °C	S-2P	0.04	17
	HexaPro	0.005	143
	Alpha S1 HexaPro	0.2	4
	S-2P + 3A3	0.1	5
37 °C→10 °C	S-2P	0.05	14
	HexaPro	0.004	171
	Alpha S1 HexaPro	0.1, 0.2 (*)	3, 5 (*)
4 °C→ 37 °C	S-2P	0.08	9
	HexaPro	0.3	2
	Alpha S1 HexaPro	0.2	4

(*) Time course was monitored twice, and the results of each fit are reported.

The prefusion conformation of spike has already been noted to be temperature dependent. Cryo-EM experiments of spike incubated at 4 °C for 5 to 7 days showed less than 10% of the definable prefusion spike particles seen on grids of freshly prepared spike. Incubating spike at 37 °C for three hours after storage at 4 °C recovered particle density to the level seen using freshly prepared protein (*30*). Failure to detect particles also correlated with a loss in recognition by an antibody known to recognize quaternary structure. These studies are consistent with our findings—long-term incubation of spike at 37 °C biases to the prefusion conformation while long-term incubation at 4 °C prefers an expanded conformation, which is apparently not well visualized on cryo-EM grids. Our results suggest that, while there is a loss of quaternary structure, the HDX protection and therefore secondary structures for each domain are similar for both states, indicating that each protomer is still structured and not denatured – a feature uniquely addressable by HDX.

3.3.4 Effects of sequence changes—HexaPro

The small energetic difference between these states indicates that small changes in sequence may affect the relative populations and/or rates of interconversion between them. Indeed, the S-2P variant was designed to stabilize the pre-fusion conformation avoiding spontaneous conversion to the post-fusion form. This S-2P construct is the basis for most currently employed vaccines. Shortly after the determination of the S-2P structure by cryo-EM, a new version was constructed, termed HexaPro or S-6P, which contains four additional proline mutations designed to increase the apparent stability of the pre-fusion state and improve cellular expression (*31*).

Using the same pulsed-labeling HDX-MS process, HexaPro shows the same bimodal behavior, with the same regions reporting on the two conformations (see Fig. 3-6A, Fig. 3-7B). At 4 °C, HexaPro, like S-2P, converts to state B, but with slower kinetics ($t_{1/2}$ of ~6 days). At 37 °C HexaPro shifts back to state A with a $t_{1/2}$ of ~2 hours (Fig. 3-6C). In sum, as expected based on the design criteria, HexaPro does result in a bias towards the prefusion conformation. Importantly, these changes demonstrate how a small number of mutations can perturb and modulate the conformational landscape of spike, suggesting that the evolving sequence variants may show differences in this conformational exchange (see below).

3.3.5 Effects of sequence changes—an interprotomer disulfide bond

To further probe the structural features of the B conformation, we turned to a variant of HexaPro engineered to contain a disulfide bond. This variant trimer contains three disulfide bonds (S383C/D985C) that reach across protomers and lock the RBDs in the down state (*32–34*). When probed by continuous-exchange HDX-MS we find that this disulfide-locked variant remains completely in the A state and does not show any observable population of the B state, even after O/N incubation at 25 °C (Fig. 3-5C). These data are consistent with a model where formation of the B state requires opening of the inter-protomer (trimer) interface and exposure of the RBDs, which would be prohibited by the interprotomer crosslinks.

3.3.6 Effects of sequence changes—B.1.1.7 (Alpha) variant

Increasingly infectious SARS-CoV-2 variants of concern are being discovered throughout the global population on a regular basis. Most of these include mutations in the spike protein, primarily in the S1 domain; some reside in the ACE2-interaction surface, while others do not. Therefore, we asked if these mutations can influence the biases and kinetics of interconversion of the A and B conformation. We monitored the A/B conversion for a variant of HexaPro that includes five S1 mutations in the B.1.1.7 (alpha) variant and none in the S2, (Δ69–70 (NTD), Δ144 (NTD), N501Y(RBD), A570D (SD1), P681H (SD2)), termed alpha S1 HexaPro. Indeed, alpha S1 HexaPro shows notable differences in both the relative preference for state B and the kinetics of interconversion. At 4 °C, alpha S1 HexaPro converts to state B nearly 20 times faster than HexaPro (Fig. 3-6C, Table 3-1). Furthermore, alpha S1 HexaPro shows no detectable prefusion conformer at 4 °C, while HexaPro shows at least 30% even after several weeks at 4 °C (Fig. 3-7B). At 37 °C, the kinetics and equilibrium distribution appear nearly identical between the two. All of the mutations are at solvent-exposed residues, except residue 570, which contacts the S2 subunit and resides in a region with observed bimodal behavior. Thus, despite their location in the S1 subunit and not at the core trimer interface, these specific B.1.1.7 mutations allosterically affect the interconversion of these two states.

3.3.7 Effects of ACE2 binding

The primary function of the RBD is to recognize the host cell receptor ACE2. In the down conformation, the RBD is occluded from binding to ACE2, and in the up conformation its accessible. The entire trimer can exist with zero, one, two, or all three RBDs in the up conformation (*7*, *15*). In the isolated RBD, the Receptor Binding Motif (RBM) should always be accessible for ACE2 binding. We used continuous-exchange HDX to monitor the binding of the receptor, both in isolation and in the full-length spike (S-2P), using a soluble dimeric form of ACE2 (ACE2-Fc, herein referred to as ACE2). For isolated RBD (residues 319–541, see Materials and Methods), we obtained 141 peptides, including one glycosylated peptide spanning the N-glycosylation site at residue 343 (no peptides are observed for site 331), resulting in 82% sequence coverage with an average redundancy of 8 (Fig. 3-2A).

The effects of ACE2 binding are illustrated in Figure 3-8. In the presence of ACE2, the latter half of the RBM (residues 472–513), shows a notable decrease in hydrogen exchange upon binding ACE2 (Fig. 3-8B), consistent with the known ACE2/RBD binding interface (*35*, *36*). We also observe small, but significant, changes for other regions that are near the binding interface. Importantly, we see very similar changes in HDX rates in RBD for both the isolated domain and in the context of the spike ectodomain, suggesting that all three RBDs in full-length spike interact with ACE2 in our experiment and that both the A and the B state can productively bind ACE2, which for the prefusion (A) state requires that RBD transition to the up state.

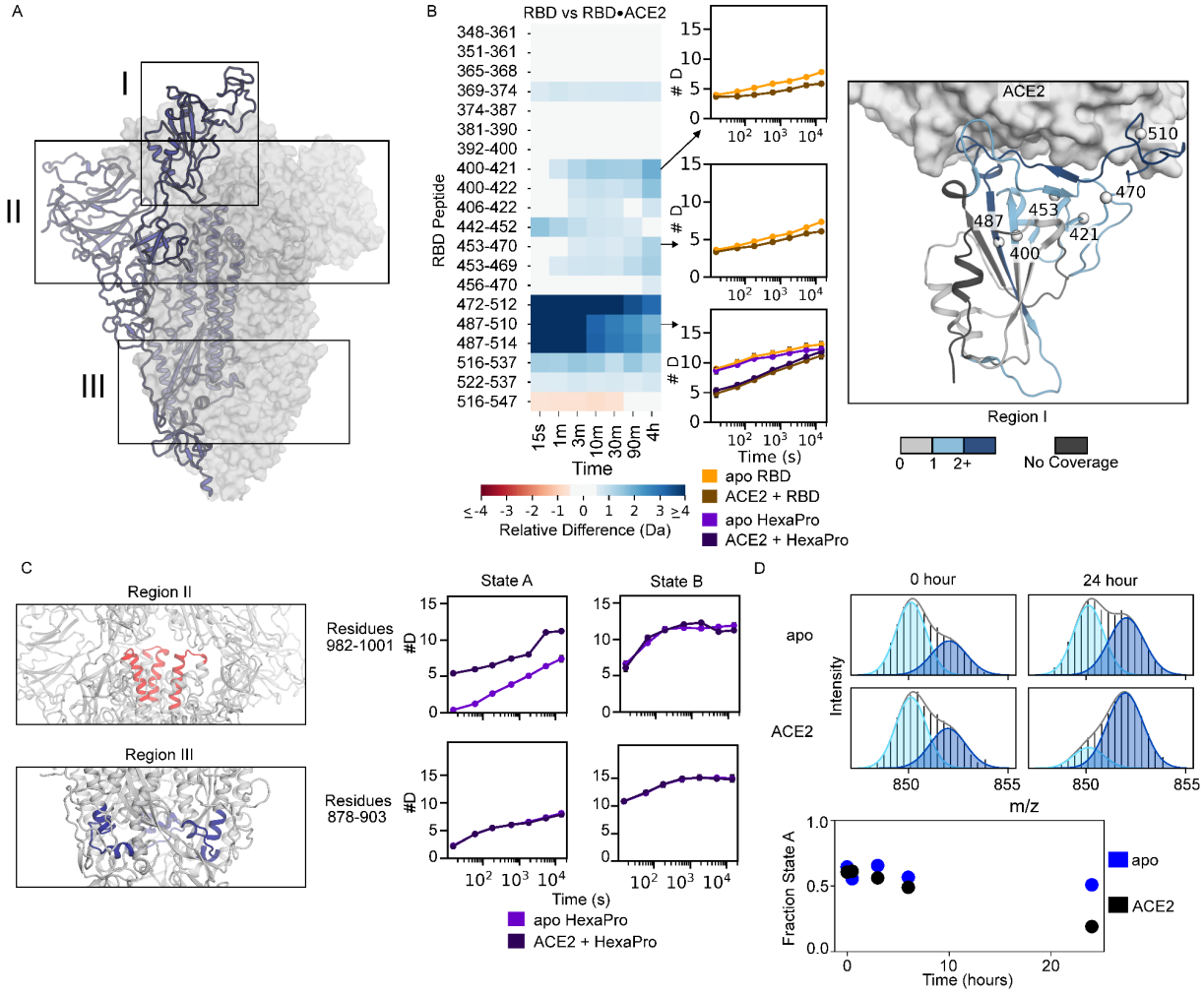


Figure 3-8. ACE2-binding effects are similar on isolated RBD and RBD in the context of the full-length ectodomain. **(A)** Diagram of the spike structure with regions of interest highlighted. **(B)** Left: Heatmap showing the difference in RBD peptide deuteration (from continuous-exchange HDX-MS) in the presence and absence of ACE2 on the isolated RBD. Middle: The deuterium uptake plots are illustrated for three peptides of interest with error bars representing the standard deviation of three replicates (some are smaller than marker size). The uptake plot for the RBM (residues 487- 510) is shown for both the isolated RBD and HexaPro. Right: Schematic representation of the heatmap data on the structure of the RBD•ACE2 complex (PDB 6M0J). The structure of the RBD is colored based on the maximum change shown in the heatmap for that residue in any peptide. **(C)** Changes to peptides from HexaPro upon binding of ACE2 outside of the RBD during continuous-exchange HDX-MS. When ACE2 binds the canonical prefusion structure, state A, peptide 982-1001 (Region II) loses inter-subunit contacts with the RBD and thus exchanges faster, but when ACE2 binds the open trimer (state B) it does not, presumably because it is already maximally exposed. For peptide 878-903 (Region III), there is no change in the exchange rate to either state A or state B indicating this region is not affected by ACE2 binding. In the schematic for region II, one NTD has been removed to visualize the peptide of interest. **(D).** Time course of interconversion in the presence of ACE2. Top: Pulsed-labeling HDX-MS example spectra of S-2P peptide 878-902 with and without ACE2 before and after 24 hours of incubation at 25 °C. Bottom: time vs fraction state A for peptide 878-902 in S-2P with and without ACE2 over 24 hours monitored by pulsed-labeling. After 24 hours ACE2 bound S-2P prefers state B.

In the context of the spike trimer, we also observe notable changes outside of the RBD, particularly in state A, where a few peptides exchange more rapidly in the presence of ACE2 (in state B these peptides do not have any notable difference in the presence of ACE2) (Fig. 3-8C). These peptides are located on the top of S2 (residues 978–1001), a region known to become more exposed when RBD transitions from a down to an up conformation. Since ACE2 binding in the prefusion state requires the RBDs to be in the up conformation, this increased exchange reflects the known biases in the RBD conformation—a prefusion state whose RBDs are primarily in the down conformation and must transition to an up conformation to bind ACE2. We also see changes in the interconversion between state A and state B in the presence of ACE2, such that state B is more preferred (Fig. 3-8D).

3.3.8 RBD dynamics are similar in isolation and in intact spike

The isolated RBD has been used for many biochemical studies and is the main component of many clinical diagnostics. It is therefore important to ask whether there are large differences in the RBD when it is in isolation versus in the context of the spike trimer; our experiments allow us to directly compare the two. Very few peptides in the RBD show substantial changes in HDX behavior (Fig. 3-9) and support the use of approaches such as deep sequence mutagenesis on the isolated RBD to gain information on the potential effects of variants, such as escape mutations (*37*).

We do, however, see some key differences in the RBD—mostly at the termini of the isolated domain and in the expected interactions with the rest of spike and across the protomer interface. The C-terminal region of the RBD (residues 516–537) is notably less protected in the isolated domain. This region is not part of the RBD globular domain and in full-length spike forms part of subdomain 1, so it is not surprising that there would be an increase in flexibility when isolated from the rest of this subdomain. Future studies with the isolated RBD may benefit from removal of both C-terminal and N-terminal (no peptide coverage observed for this region) regions, as they are likely disordered and may interfere with crystallization or lead to increases in aggregation.

3.3.9 3A3—an antibody that binds specifically to the B state

Recently, an antibody, 3A3, was developed that binds to MERS-CoV, SARS-CoV-1, and SARS-CoV-2, with an apparent epitope in a region where we observe bimodal behavior (residues ~980–1000) (*28*). This region, however, is inaccessible in the prefusion structure—it is buried in the prefusion structure when all RBDs are down, and highly occluded when the RBDs are up. Our HDX data indicate that this region is exposed in state B. To confirm the epitope, we repeated the continuous-exchange HDX studies in the presence of 3A3; indeed, we see strong increased protection in the 978–1001 region. Moreover, this protection is directly associated with state B as evidenced by the now unimodal distribution for peptides in region 978-1001 (Fig. 3-10A). This unimodal distribution can be explained by a decrease in hydrogen exchange in state B due to direct binding of 3A3 to this region such that the exchange is similar in both the A and B state. These data support a model where the 3A3 antibody binds uniquely to the B state.

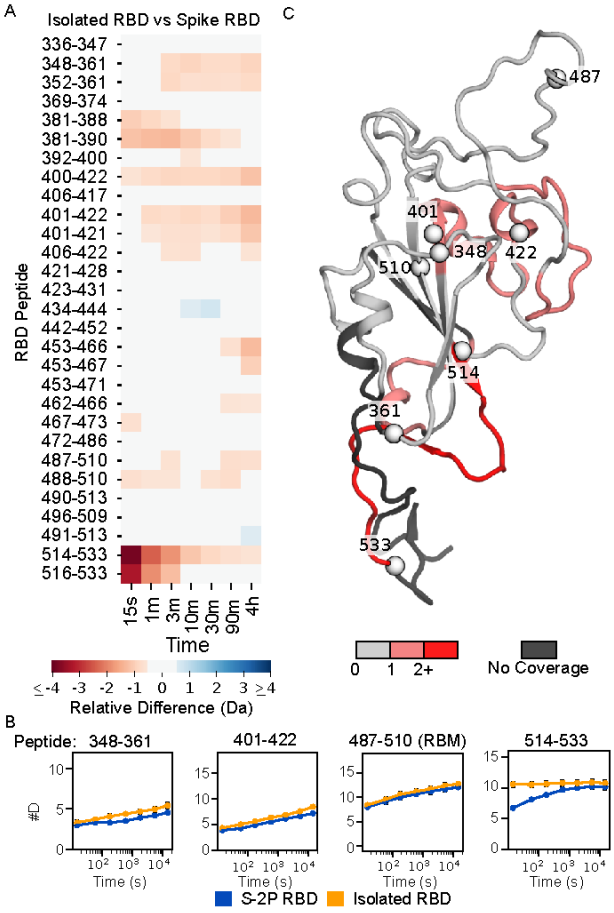


Figure 3-9. Comparison of HDX on RBD in isolation versus in S-2P

Left: Heat map showing the difference in peptide deuteration for isolated RBD compared to the RBD in S-2P. Bottom: Selected uptake plots of isolated RBD and S-2P RBD. Right: Structure of the RBD (model of a single RBD taken from a full-length spike trimer model from *(24)*) colored based on the maximum change shown in the heat map for that residue in any peptide. For reference, spheres are shown denoting the beginning and end of the peptides displayed in the uptake plots.

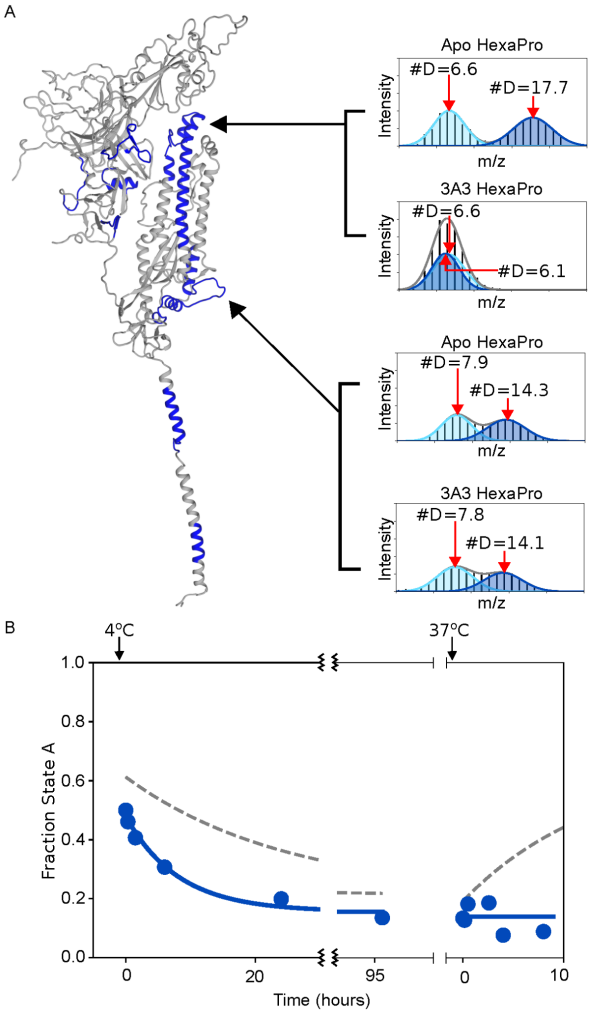


Figure 3-10. The antibody 3A3 binds selectively to state B in the 978-1001 region. (A) Example mass spectra for HexaPro with and without 3A3 for two different peptides that have bimodal mass distributions. The bottom peptide (878-902) shows no change in the presence of 3A3, which indicates that the amount of state A and state B has not significantly changed 13 minutes after adding 3A3. The top peptide (978-1001), however, shows significant protection in the presence of 3A3, shifting the distribution belonging to state B to a deuteration amount indistinguishable from state A. These data are the three-minute time point from a continuous-exchange HDX-MS time course. **(B)** The kinetics of interconversion of S-2P in the presence of 3A3 monitored by pulsed-labeling HDX-MS. The addition of 3A3 accelerated the rate of conversion to state B at 4 °C. The binding of 3A3 prevents the return to state A at 37 °C. Dotted lines indicate the conversion in the absence of 3A3.

To confirm this hypothesis, we looked at the effect of 3A3 binding on the temperature-induced conversion between A and B using our pulsed-labeling HDX-MS method. 3A3 increases the rate of conversion from A to B at 4 °C, decreasing the $t_{1/2}$ from ~17 hours to ~5 hours (Fig. 3-10B, Table 3-1). This increase in the observed rate implies that 3A3 also affects the transition state for the conversion. Furthermore, returning the sample to 37 °C in the presence of 3A3 (state B saturated with 3A3) prohibits any transition back to the prefusion state, indicating that the binding of 3A3 prevents formation of the prefusion state most likely due to steric hindrance of the antibody being bound to the trimer interface. Since 3A3 binds wild-type and D614G spike when expressed on the surface of cells and neutralizes pseudovirus expressing these spikes (*28*), the data collectively suggest state B exposes broadly neutralization-sensitive epitopes that may be of interest for future therapeutics and vaccines.

3.4 Discussion

The above data allow us to create a structural model for state B (Fig. 3-11). The overall fold, or topology, of each domain is likely similar to the prefusion structure as, with the exception of the bimodal peptides, their hydrogen exchange patterns are similar. The bimodal peptides, which report on the two different conformations, cluster in the trimer interface, suggesting that this interface is less protected in state B. State B is not a monomer. Size-exclusion chromatography combined with multi-angle light scattering (SEC-MALS) and the hydrogen exchange data at the trimerization motif, confirm that both conformations are trimeric (Fig. 3-12). In these soluble ectodomain constructs, the trimer is held together by the appended C-terminal trimerization domain, while in the full-length native spike trimer, the transmembrane helical segment likely serves this function. Therefore, state B is best modeled as an opened-up trimer with three protomers whose domains are structurally uncoupled. Our data do not let us address the relative orientation of the protomers within individual trimers; however, an ensemble of opened-up trimers with heterogeneous positioning of the protomers would best explain the lack of cryo-EM data. An opened-up class 1 viral fusion protein has been reported for respiratory syncytial virus (RSV) and visualized by a low resolution structure (*38*). This structural data from RSV and reports of an opening up of other viral fusion proteins (*39–41*) support our model of an ensemble of open-trimers with various degrees of openness.

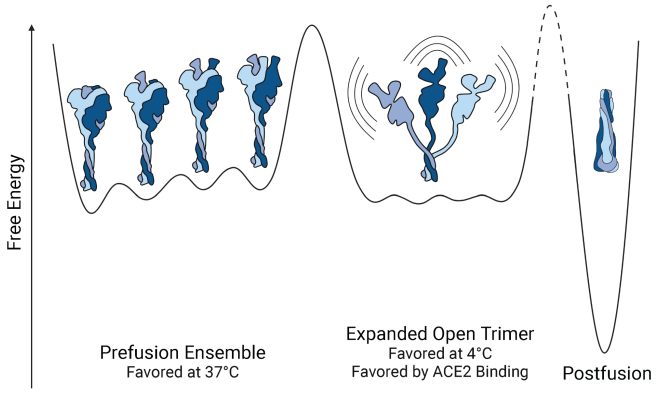


Figure 3-11. Schematic of the energy landscape for the SARS-CoV-2 Spike ectodomain. Reaction coordinate illustrating the different conformations accessible by Spike. Three different conformational states are depicted: the canonical prefusion ensemble, the expanded open trimer, and the postfusion conformation. The prefusion conformation contains all four RBD states (0,1,2, or 3 up). The relative energies and barrier heights as well as the placement of the open trimer along the reaction coordinate are drawn for illustration only.

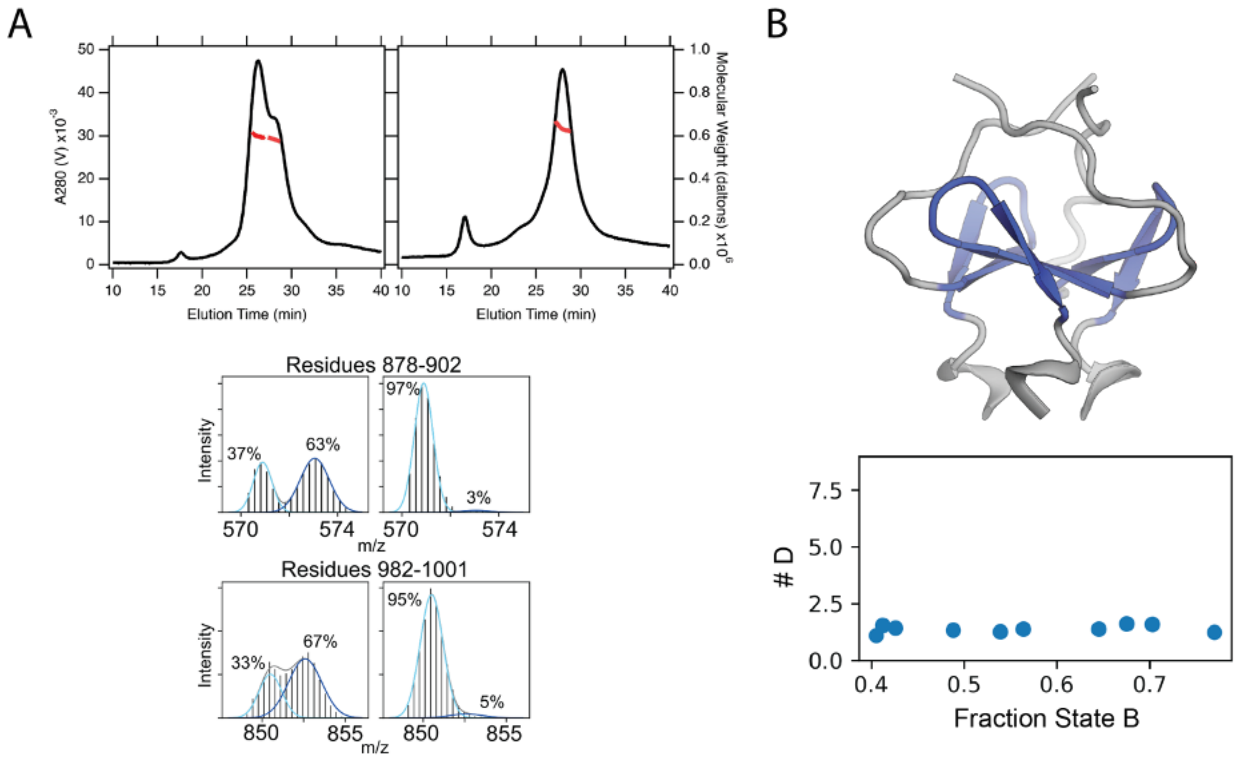


Figure 3-12. State A and state B are both Trimeric. (A) Top: SEC-MALS (Superose 6 increase 10/300) UV (black line, left axis) and estimated molecular weight (red line, right axis) traces from S-2P performed after four days of incubation at 4 °C (left) and 37 °C (right). The molecular weights of all primary species are most consistent with the spike ectodomain trimer (~600 kilodaltons). Bottom: MS spectra of a bimodal peptides from each sample taken immediately before SEC-MALS experiment with gaussian fits and estimated populations shown. Populations of states A and B are consistent with the observed peaks shapes in the SEC-MALS experiments (one predominant peak at 37 °C, and two predominant peaks with a ratio of ~3:2 at 4 °C) the earlier elution volume of state B supports a model of a more expanded conformation (Fig. 3-7) (B) Top: Structure of the T4 fibritin trimerization domain (PDB 1RFO) with peptide shown in blue. Bottom: peptide deuterium uptake at one minute as a function of fraction state B. The observed lack of dependence is consistent with the maintenance of the trimerization domain structure in both states A and B.

We propose that this conformational change occurs because the trimer interface of the spike protein consists of weak, hydrophobic interactions which are further weakened at low temperatures. This can be thought of as a form of cold denaturation where the open trimer has a higher heat capacity than the canonical prefusion structure, consistent with solvation of the hydrophobic trimer interface. As for what is responsible for the slow rate of interconversion, this is still unknown. We believe that in order for state B to be formed all three RBDs would need to be in the up position, which could be the rate-limiting step for the formation of state B, and thus the slow rate may be a result of the intertwined nature of the trimer. While it is known that RBDs moving up and down occurs on the order of seconds, having all three RBDs being open simultaneously long enough for the trimer interface to disassociate, may be rare enough to result

in a high kinetic barrier. Similarly, the role of the designed prolines (in S-2P and HexaPro) is unclear. The rates of interconversion are not consistent with a model where the isomerization of prolines is playing a role. Rather, our data are consistent with the model where the introduction of prolines successfully stabilizes the prefusion conformation.

A loss of interprotomer contacts in state B implies that the RBDs no longer contact adjacent protomers, and thus do occupy distinct 'down' and 'up' conformations; rather, they are likely always in a binding-competent state, perhaps even more accessible than the canonical 'up' state. This increased availability of the RBD may drive a preference for the B state in the presence of ACE2. Furthermore, in the prefusion conformation, having all three RBDs bound to ACE2 could lead to steric hindrances, but in state B, all three RBDs should be able to bind ACE2 with high affinity. Interestingly, introduction of variants of concern, such as in the alpha S1 HexaPro variant greatly increases the rate of conversion to state B. Whether this plays a role in the noted increased infectivity remains to be determined.

Molecular dynamics has shown a smaller opening of the spike protein where an RBD and adjacent NTD twist and peel away from the center of the spike protein, revealing a cryptic epitope at the top of the S2 domain (*20*). This rapidly sampled conformation is not state B, as it does not involve the S2 trimer interface and the timescale of conversion to state B is unlikely to be sampled during a molecular dynamics simulation. This partial opening, however, could be on pathway to state B.

3.4.1 A potential role for state B in spike function

The increase in the formation of state B upon binding ACE2 suggests that state B may be a functional intermediate. While we cannot say whether this transition occurs in wild-type spike on virions, it is possible that state B represent an intermediate along the pathway to S1 shedding during the transition from the prefusion conformation to the postfusion conformation. This irreversible transition is not possible in the soluble ectodomain which removes the proteolytic cleavage site. If this is an on-pathway intermediate, antibodies or ligands that trap the protein in this state may block the protein along the pathway to fusion (*42*) those that act on the transition state and increase the rate of formation of the intermediate could promote the premature formation of the postfusion conformation and thus aid in neutralization. If instead, formation of state B is off-pathway, antibodies or other ligands that favor state B would essentially trap the protein in an inactive conformation.

3.4.2 The alternative conformation presents new druggable sites

Based on our model, the newly identified B state contains a large and unique accessible surface area that is buried in the canonical prefusion conformation, thus exposing different epitopes for antibody and ligand recognition. Moreover, these regions arise from the most highly conserved part of the protein, the S2 trimer interface, and therefore may present an ideal target for vaccines that would provide protection across coronaviruses (pan-coronavirus). In terms of therapeutics, ligands directed towards this region may also be broadly efficacious against variants of concern

as well as other coronaviruses. Indeed, the antibody 3A3 represents one such potential therapeutic.

3.4.3 State B will affect measured ligand-binding affinities.

Finally, independent of whether this additional conformation is an on-pathway intermediate in the coronavirus functional lifecycle, it is ubiquitous among *in vitro* preparations of the spike protein. We see evidence of this conformation in every variant examined, excluding the disulfide-locked sample. Many biochemical and diagnostic assays use these isolated spike constructs, and many laboratories store this protein at 4 °C, where the alternative, expanded conformation is favored. Given that state A and B have differing affinities for the receptor and some antibodies, the temperature and time-dependent changes in the population of state B complicates quantitative analysis of binding affinities and needs to be further evaluated.

3.5 Conclusion

In sum, we have found that the SARS-CoV-2 spike ectodomain reversibly samples an open-trimer conformation, potentially allowing for the development of pan-coronavirus vaccines and therapeutics. This open trimer is folded and exposes a highly conserved region of the protein. It is similar in energy to the well-characterized prefusion conformation determined by cryo-EM. The fraction of spike in each conformer depends on temperature, ligands, and sequence. Mutations, receptor binding, antibodies, and temperature all affect the kinetics and energetics of this conformational change. Thus, quantitative measurements, such as *in vitro* binding assays, need to be re-evaluated for possible effects due to this mixed population. How easily this conformation is sampled in natural membrane-bound spike and its position in the viral lifecycle are still unknown; however, an antibody specific for this conformation can bind and neutralize in *in vitro* SARS-CoV-2 cell fusion and pseudovirus assays suggesting an important role. Determining which conformation elicits more robust protective immunity against both SARS-CoV-2 variants and other coronaviruses will be important for future vaccine development.

3.6 Methods

3.6.1 Protein expression and purification

SARS-CoV-2 Spike (2P) and RBD were expressed and purified from stably transformed Expi293 cells, following methods as described (*43*). Briefly, stable Expi293 suspension cells were maintained in Expi293 media at 37°C. Cells were grown for 6 days, then harvested by centrifugation, and filtered. Supernatant was adjusted to pH 7.4 and loaded onto a 5 mL HisTrap Excel column and washed with 60 CV of buffer. Captured proteins were eluted with 10 CV of (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4). Eluted proteins were buffer exchanged into PBS using either 3kDa MWCO (for RBD) or 100 kDa MWCO (for Spike) Amicon concentrators and filtered prior to storage at -80°C. HexaPro, HexaPro S383C/D985C, and alpha S1 HexaPro were expressed and purified from transiently transfected Freestyle 293-F

cells as described (*31*). Briefly, cell cultures were harvested four days after transfection and centrifuged. Supernatants were filtered and flowed over StrepTactin resin (IBA). Proteins were further purified by size-exclusion chromatography using a Superose 6 increase column (GE Healthcare) in a buffer composed of 2 mM Tris pH 8.0, 200 mM NaCl and 0.02% NaN3. ACE2-Fc was a gift from the Wells lab (*44*). 3A3 IgG was expressed and purified from ExpiCHO cells as described (*28*). Briefly, Antibodies were expressed in ExpiCHO cells according to the high titer protocol provided and purified on a Protein A HiTrap column (GE Healthcare) and buffer exchanged to PBS.

3.6.2 Continuous hydrogen exchange labeling

For all continuous hydrogen exchange experiments, deuterated buffer was prepared by lyophilizing PBS (pH 7.4, Sigma-Aldrich P4417) and resuspending in D_2O (Sigma-Aldrich 151882). To initiate the continuous exchange experiment, samples were diluted 10-fold (final spike trimer concentration of 0.167 μM) into temperature equilibrated deuterated PBS buffer (pH_{read} 7, pD 7.4). Samples were quenched, at the timepoints outlined below, by mixing 30 μL of the partially exchanged protein with 30 μL of 2x quench buffer (3.6 M GdmCl, 500 mM TCEP, 200 mM Glycine pH 2.4) on ice. Samples were incubated on ice for one minute to allow for partial unfolding to assist with proteolytic degradation and then flash frozen in liquid nitrogen and stored at -80°C.

For studies comparing HexaPro ± ACE2 the RBD in isolation vs in S-2P, purified spike (1.67 μM spike trimer or 5 μM RBD) was incubated in PBS at 25°C overnight (12-16 hours) before the initiation of hydrogen exchange. For experiments done in the presence of ACE2-Fc, the ligand was added during this incubation at a 1.25:1 molar ratio of ligand to spike monomer (6.25 μM ligand) to ensure saturation. Based on the reported affinity (K_D ~15 nM) for ACE2-Fc, fraction bound can be assumed to be greater than 97%. The hydrogen exchange time points for these experiments were 15 seconds, 60 seconds, 180 seconds, 600 seconds, 1800 seconds, 5400 seconds, and 14400 seconds.

For the comparison of HexaPro ± 3A3, HexaPro was incubated overnight at 37 °C (12-16 hours). After incubation the protein was moved to 25 °C and diluted to 1.67 μM spike trimer. In the 3A3 bound condition, 6.25 μM antibody was added and allowed to bind for 10 minutes at 25 °C. Given the affinity of 3A3 for HexaPro (12 nM, fraction bound can be assumed to be greater than 97%. The quench time points for this experiment were 15 seconds, 180 seconds, 1800 seconds and 14400 seconds.

3.6.3 Back exchange control preparation

S-2P was diluted to 1.67 μM trimer in PBS pH 7.4. To initiate hydrogen exchange, the sample was diluted 10-fold (final spike trimer concentration of 0.167 μM) into deuterated PBS buffer (pH_{read} 7, pD 7.4) that was supplemented with 3.6 M GdmCl, and then incubated at 37 °C. The addition of denaturant and increased temperature should both promote hydrogen exchange, by destabilizing folded structures and increasing the intrinsic rate of hydrogen exchange,

respectively. Following two weeks of exchange, 30 μL of deuterated spike was mixed with 30 μL of 2X quench buffer lacking denaturant (500 mM TCEP, 200 mM Glycine pH 2.4) and kept on ice for one minute prior to flash freezing in liquid nitrogen and storage at -80 °C. The results of this control experiment were used to characterize the back exchange of the system and were not used to adjust deuteration values of continuous-exchange experiments.

3.6.4 Incubation kinetics and pulse labeling

For evaluating the temperature dependent kinetics of interconversion, frozen spike samples were thawed, diluted to 5 μM spike monomer, and incubated at 37°C overnight. Samples were then moved to a temperature-controlled chamber at 4°C and the population of each state was evaluated at the specified time points as described below. After the final 4 °C sample was taken (96–526 hours, depending on the spike construct), the sample was returned to a 37° heat block for further incubation and the population of each state was again evaluated at the specified time points as described below. To evaluate the relative population of the A and B conformer at each time point, 3 μL of spike sample was removed from the incubation tube and mixed with 27 μL of room temperature deuterated buffer. After a one-minute labeling pulse, 30 μL of quench buffer kept on ice was mixed with the 30 μL of labeled protein. Quenched samples were kept on ice for one minute to allow for partial unfolding, and then flash frozen in liquid nitrogen.

For kinetics carried out in the presence of ACE2 or 3A3, after the initial 37 °C incubation the sample was brought to 25 °C and ligand was added (6.25 μM). To monitor the population of state A and B as a function of time at 25 °C in the presence of ACE2 the sample was kept in a temperature-controlled chamber at 25 °C and aliquots were removed for pulse labeling as described above at 0 hours, 30 minutes, 3 hours, 6 hours and 24 hours.

3.6.5 Protease digestion and LC/MS

All samples were thawed immediately before injection into a cooled valve system (Trajan LEAP) coupled to a LC (Thermo UltiMate 3000). Sample time points were injected in random order. The temperature of the valve chamber, trap column, and analytical column were maintained at 2 °C. The temperature of the protease column was maintained at 10 °C. The quenched sample was subjected to inline digestion by two immobilized acid proteases in order, aspergillopepsin (Sigma-Aldrich P2143) and porcine pepsin (Sigma-Aldrich P6887) at a flow rate of 200 μL/min of buffer A (0.1 % formic acid). Protease columns were prepared in house by coupling protease to beads (Thermo Scientific POROS 20 Al aldehyde activated resin 1602906) and packed by hand into a column (2mm ID x 2 cm, IDEX C-130B). Following digestion, peptides were desalted for 4 minutes on a hand-packed trap column (Thermo Scientific POROS R2 reversed-phase resin 1112906, 1 mm ID x 2 cm, IDEX C-128). Peptides were then separated with a C8 analytical column (Thermo Scientific BioBasic-8 5 μm particle size 0.5 mm ID x 50 mm 72205-050565) and a gradient of 5-40% buffer B (100% Acetonitrile, 0.1% Formic Acid) at a flow rate of 40 μL/min over 14 minutes, and then of 40-90% buffer B over 30 seconds. The analytical and trap columns were then subjected to a sawtooth wash and equilibrated at 5% buffer B prior to the next injection. Protease columns were washed with two injections of 100 μL 1.6 M GdmCl, 0.1% formic acid prior to the next injection. Peptides were eluted directly into a Q

Exactive Orbitrap Mass Spectrometer operating in positive mode (resolution 70000, AGC target 3e6, maximum IT 50 ms, scan range 300-1500 m/z). For each spike construct, a tandem mass spectrometry experiment was performed (Full MS settings the same as above, dd-MS^2 settings as follows: resolution 17500, AGC target 2e5, maximum IT 100 ms, loop count 10, isolation window 2.0 m/z, NCE 28, charge state 1 and ≥7 excluded, dynamic exclusion of 15 seconds) on undeuterated samples. LC and MS methods were run using Xcalibur 4.1 (Thermo Scientific)

3.6.6 Peptide identification

Byonic (Protein Metrics) was used to identify unmodified and glycosylated peptides in the tandem mass spectrometry data. The sequence of the expressed construct, including signal sequence and trimerization domain, was used as the search library. Sample digestion parameters were set to non-specific. Precursor mass tolerance and fragment mass tolerance was set to 6 and 10 ppm respectively. Variable N-linked glycosylation was allowed, with a library of 132 human N-glycans used in the search. No non-glycosylated peptides spanning any of the 22 known glycosylation sites in the spike sequence were ever observed, independent of the glycosylation search parameters. Peptide lists (sequence, charge state, and retention time) were exported from Byonic and imported into HDExaminer 3 (Sierra Analytics). When multiple peptide lists were obtained, all were imported and combined in HDExaminer 3.

3.6.7 HDExaminer 3 analysis

Peptide isotope distributions at each exchange time point were fit in HDExaminer 3. For glycosylated peptides, only the highest confidence modification was included in the mass spectra search and analysis. For unimodal peptides, deuteration levels were determined by subtracting mass centroids of deuterated peptides from undeuterated peptides. For bimodal peaks, extracted peptide isotope spectra were exported from HDExaminer 3 and analyzed separately (see below for details).

3.6.8 Bimodal fitting and conformation quantification

Peptide mass spectra for bimodal peptides were exported from HDExaminer 3.0. All quantitative analysis of the exported peptide mass spectra was performed using python scripts in Jupyter notebooks. After importing a peptide mass spectra, the m/z range containing all possible deuteration states of the selected peptide was isolated and the find_peaks method from the scipy.signal package was used to identify each isotope in the mass envelope and the height of each peak was used as its intensity. The area of the total mass envelope was normalized to account for run-to-run differences in intensity. The bimodal mass envelopes for all timepoints under the same condition were globally fit to a sum of two gaussians, keeping the center and width of each gaussian constant across all incubation time points. Fitting was done using the curve_fit function from the scipy.optimize package. After fitting, the area under each individual gaussian was determined to approximate the relative population of each state.

3.6.9 Size Exclusion Chromatography – Multiangle Light Scattering (SEC-MALS)

To generate variable populations of states A and B, S-2P (100 µL, 0.18 mg/mL) was incubated at either 4 °C or 37 °C for four days. Post-incubation, and prior to SEC-MALS injection, an aliquot from each sample was taken and a one-minute pulse-labeling experiment was performed (see above). The resulting bimodal peptide distributions were used to calculate a fraction state A and B. Samples were filtered (0.22 um
um hydrophilic polyvinylidene fluoride Ultrafree-MC GV centrifugal filter) prior to SEC-MALS injection. 90 µL of sample (0.18 mg/mL) was then injected onto a Superose 6 increase 10/300 (GE Healthcare) preequilibrated in filtered (0.1 um polyethersulfone Nalgene Rapid-Flow vacuum filter) PBS flowing at 0.5 ml/min at 4 °C. The ÄKTA Pure 25 M1 (Cytiva) chromatography system was coupled to an 18-angle light scattering Wyatt Dawn detector and Wyatt Optilab refractive index detector (Wyatt Technology). Data analysis was carried out using the program ASTRA 7.1.4.8.

3.7 Acknowledgements

We thank the entire Marqusee Lab for advice and critical reading of the manuscript, particularly Johanna Lindner for help initiating these experiments. We thank Peter S. Kim, Abigail Powell, Payton Weidenbacher, and Natalia Freedland (Stanford University) for advice and discussion; Jim Wells, Shion Lim, and Irene Lui (University of California, San Francisco) for discussion and providing ACE2-Fc; Yifan Cheng, Adamo Mancino, and Mingliang Jin (University of California, San Francisco) for assistance with SEC-MALS and Aashish Manglik (University of California, San Francisco) for advice and discussion. We thank Jeff Morrow and Sierra Analytics for making HDExaminer freely available for at-home use during shelter-in-place orders.

SM is a Chan Zuckerberg Biohub Investigator. Some figures were created with the use of BioRender.com. This work was funded by NSF Graduate Research Fellowship DGE1752814 (SMC); NIH grant no. GM050945 (SM); and NSF grant no. MCB1616591 (SM); NIH grant no. AI127521 (JSM); Welch foundation grant no. F-0003-19620604 (JSM); NIH grant no. AI122753 (JAM); Welch foundation grant no. F-1767 (JAM). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

3.8 References

1. A. Baum, B. O. Fulton, E. Wloga, R. Copin, K. E. Pascal, V. Russo, S. Giordano, K. Lanza, N. Negron, M. Ni, Y. Wei, G. S. Atwal, A. J. Murphy, N. Stahl, G. D. Yancopoulos, C. A. Kyratsous, Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*. **369**, 1014–1018 (2020).

2. P. C. Taylor, A. C. Adams, M. M. Hufford, I. de la Torre, K. Winthrop, R. L. Gottlieb, Neutralizing monoclonal antibodies for treatment of COVID-19. *Nat. Rev. Immunol.* **21**, 382–393 (2021).

3. J. Hansen, A. Baum, K. E. Pascal, V. Russo, S. Giordano, E. Wloga, B. O. Fulton, Y. Yan, K. Koon, K. Patel, K. M. Chung, A. Hermann, E. Ullman, J. Cruz, A. Rafique, T. Huang, J. Fairhurst, C. Libertiny, M. Malbec, W.-Y. Lee, R. Welsh, G. Farr, S. Pennington, D. Deshpande, J. Cheng, A. Watty, P. Bouffard, R. Babb, N. Levenkova, C. Chen, B. Zhang, A. Romero Hernandez, K. Saotome, Y. Zhou, M. Franklin, S. Sivapalasingam, D. C. Lye, S. Weston, J. Logue, R. Haupt, M. Frieman, G. Chen, W. Olson, A. J. Murphy, N. Stahl, G. D. Yancopoulos, C. A. Kyratsous, Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science*. **369**, 1010–1014 (2020).

4. R. Shi, C. Shan, X. Duan, Z. Chen, P. Liu, J. Song, T. Song, X. Bi, C. Han, L. Wu, G. Gao, X. Hu, Y. Zhang, Z. Tong, W. Huang, W. J. Liu, G. Wu, B. Zhang, L. Wang, J. Qi, H. Feng, F.-S. Wang, Q. Wang, G. F. Gao, Z. Yuan, J. Yan, A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature*. **584**, 120–124 (2020).

5. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. **91**, eabb9983 (2020).

6. M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. **181**, 271-280.e8 (2020).

7. D. J. Benton, A. G. Wrobel, P. Xu, C. Roustan, S. R. Martin, P. B. Rosenthal, J. J. Skehel, S. J. Gamblin, Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature*. **588**, 327–330 (2020).

8. S. Jiang, C. Hillyer, L. Du, Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. *Trends Immunol.* **41**, 355–359 (2020).

9. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. **367**, 1260–1263 (2020).

10. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veesler, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. **181**, 281-292.e6 (2020).

11. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

12. J. M. White, S. E. Delos, M. Brecher, K. Schornberg, Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit. Rev. Biochem. Mol. Biol.* **43**, 189–219 (2008).

13. F. Li, Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol*. **3**, 237–261 (2016).

14. T. Zhou, Y. Tsybovsky, J. Gorman, M. Rapp, G. Cerutti, G.-Y. Chuang, P. S. Katsamba, J. M. Sampson, A. Schön, J. Bimela, J. C. Boyington, A. Nazzari, A. S. Olia, W. Shi, M. Sastry, T. Stephens, J. Stuckey, I.-T. Teng, P. Wang, S. Wang, B. Zhang, R. A. Friesner, D. D. Ho, J. R. Mascola, L. Shapiro, P. D. Kwong, Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe*. **28**, 867-879.e5 (2020).

15. T. Xiao, J. Lu, J. Zhang, R. I. Johnson, L. G. A. McKay, N. Storm, C. L. Lavine, H. Peng, Y. Cai, S. Rits-Volloch, S. Lu, B. D. Quinlan, M. Farzan, M. S. Seaman, A. Griffiths, B. Chen, A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent. *Nat. Struct. Mol. Biol.* **28**, 202–209 (2021).

16. S. Belouzard, V. C. Chu, G. R. Whittaker, Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5871–5876 (2009).

17. Y. Cai, J. Zhang, T. Xiao, H. Peng, S. M. Sterling, R. M. Walsh Jr, S. Rawson, S. Rits-Volloch, B. Chen, Distinct conformational states of SARS-CoV-2 spike protein. *Science*. **369**, 1586–1592 (2020).

18. A. C. Walls, M. A. Tortorici, J. Snijder, X. Xiong, B.-J. Bosch, F. A. Rey, D. Veesler, Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11157–11162 (2017).

19. M. Lu, P. D. Uchil, W. Li, D. Zheng, D. S. Terry, J. Gorman, W. Shi, B. Zhang, T. Zhou, S. Ding, R. Gasser, J. Prévost, G. Beaudoin-Bussières, S. P. Anand, A. Laumaea, J. R. Grover, L. Liu, D. D. Ho, J. R. Mascola, A. Finzi, P. D. Kwong, S. C. Blanchard, W. Mothes, Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles. *Cell Host Microbe*. **28**, 880-891.e8 (2020).

20. M. I. Zimmerman, J. R. Porter, M. D. Ward, S. Singh, N. Vithani, A. Meller, U. L. Mallimadugula, C. E. Kuhn, J. H. Borowsky, R. P. Wiewiora, M. F. D. Hurley, A. M.

Harbison, C. A. Fogarty, J. E. Coffland, E. Fadda, V. A. Voelz, J. D. Chodera, G. R. Bowman, SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* **13**, 651–659 (2021).

21. T. Sztain, S.-H. Ahn, A. T. Bogetti, L. Casalino, J. A. Goldsmith, E. Seitz, R. S. McCool, F. L. Kearns, F. Acosta-Reyes, S. Maji, G. Mashayekhi, J. A. McCammon, A. Ourmazd, J. Frank, J. S. McLellan, L. T. Chong, R. E. Amaro, A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.*, 1–6 (2021).

22. S. W. Englander, Hydrogen exchange and mass spectrometry: A historical perspective. *J. Am. Soc. Mass Spectrom.* **17**, 1481–1489 (2006).

23. J. Zheng, T. Strutzenberg, B. D. Pascal, P. R. Griffin, Protein dynamics and conformational changes explored by hydrogen/deuterium exchange mass spectrometry. *Curr. Opin. Struct. Biol.* **58**, 305–313 (2019).

24. L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, R. E. Amaro, Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).

25. B. Turoňová, M. Sikora, C. Schürmann, W. J. H. Hagen, S. Welsch, F. E. C. Blanc, S. von Bülow, M. Gecht, K. Bagola, C. Hörner, G. van Zandbergen, S. Mosalaganti, A. Schwarz, R. Covino, M. D. Mühlebach, G. Hummer, J. K. Locker, M. Beck, In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges *Science*. **370**, 203-208 (2020).

26. C. O. Barnes, A. P. West Jr, K. E. Huey-Tubman, M. A. G. Hoffmann, N. G. Sharaf, P. R. Hoffman, N. Koranda, H. B. Gristick, C. Gaebler, F. Muecksch, J. C. C. Lorenzi, S. Finkin, T. Hägglöf, A. Hurley, K. G. Millard, Y. Weisblum, F. Schmidt, T. Hatziioannou, P. D. Bieniasz, M. Caskey, D. F. Robbiani, M. C. Nussenzweig, P. J. Bjorkman, Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell*. **182**, 828-842.e16 (2020).

27. P. V. Raghuvamsi, N. K. Tulsian, F. Samsudin, X. Qian, K. Purushotorman, G. Yue, M. M. Kozma, W. Y. Hwa, J. Lescar, P. J. Bond, P. A. MacAry, G. S. Anand, SARS-CoV-2 S protein:ACE2 interaction reveals novel allosteric targets. *Elife*. **10** (2021), doi:10.7554/eLife.63646.

28. Y. Huang, A. W. Nguyen, C.-L. Hsieh, R. Silva, O. S. Olaluwoye, R. E. Wilen, T. S. Kaoud, L. R. Azouz, A. N. Qerqez, K. C. Le, A. L. Bohanon, A. M. DiVenere, Y. Liu, A. G. Lee, D. Amengor, K. N. Dalby, S. D'Arcy, J. S. McLellan, J. A. Maynard, Identification of a conserved neutralizing epitope present on spike proteins from all highly pathogenic coronaviruses. *bioRxiv* (2021), p. 2021.01.31.428824, , doi:10.1101/2021.01.31.428824.

29. M. A. Benhaim, V. Mangala Prasad, N. K. Garcia, M. Guttman, K. K. Lee, Structural monitoring of a transient intermediate in the hemagglutinin fusion machinery on influenza virions. *Sci. Adv.* **6**, eaaz8822 (05/2020).

30. R. J. Edwards, K. Mansouri, V. Stalls, K. Manne, B. Watts, R. Parks, K. Janowska, S. M. C. Gobeil, M. Kopp, D. Li, X. Lu, Z. Mu, M. Deyton, T. H. Oguin 3rd, J. Sprenz, W. Williams, K. O. Saunders, D. Montefiori, G. D. Sempowski, R. Henderson, S. Munir Alam, B. F. Haynes, P. Acharya, Cold sensitivity of the SARS-CoV-2 spike ectodomain. *Nat. Struct. Mol. Biol.* **28**, 128–131 (2021).

31. C.-L. Hsieh, J. A. Goldsmith, J. M. Schaub, A. M. DiVenere, H.-C. Kuo, K. Javanmardi, K. C. Le, D. Wrapp, A. G. Lee, Y. Liu, C.-W. Chou, P. O. Byrne, C. K. Hjorth, N. V. Johnson, J. Ludes-Meyers, A. W. Nguyen, J. Park, N. Wang, D. Amengor, J. J. Lavinder, G. C. Ippolito, J. A. Maynard, I. J. Finkelstein, J. S. McLellan, Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*. **369**, 1501–1505 (2020).

32. R. Henderson, R. J. Edwards, K. Mansouri, K. Janowska, V. Stalls, S. M. C. Gobeil, M. Kopp, D. Li, R. Parks, A. L. Hsu, M. J. Borgnia, B. F. Haynes, P. Acharya, Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020).

33. M. McCallum, A. C. Walls, J. E. Bowen, D. Corti, D. Veesler, Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat. Struct. Mol. Biol.* (2020), doi:10.1038/s41594-020-0483-8.

34. The CITIID-NIHR COVID-19 BioResource Collaboration, X. Xiong, K. Qu, K. A. Ciazynska, M. Hosmillo, A. P. Carter, S. Ebrahimi, Z. Ke, S. H. W. Scheres, L. Bergamaschi, G. L. Grice, Y. Zhang, J. A. Nathan, S. Baker, L. C. James, H. E. Baxendale, I. Goodfellow, R. Doffinger, J. A. G. Briggs, A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat. Struct. Mol. Biol.* (2020), doi:10.1038/s41594-020-0478-5.

35. R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. **367**, 1444–1448 (2020).

36. Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K.-Y. Yuen, Q. Wang, H. Zhou, J. Yan, J. Qi, Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*. **181**, 894-904.e9 (2020).

37. T. N. Starr, A. J. Greaney, A. Addetia, W. W. Hannon, M. C. Choudhary, A. S. Dingens, J. Z. Li, J. D. Bloom, Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science*. **371**, 850–854 (2021).

38. M. S. A. Gilman, P. Furmanova-Hollenstein, G. Pascual, A. B van 't Wout, J. P. M. Langedijk, J. S. McLellan, Transient opening of trimeric prefusion RSV F proteins. *Nat. Commun.* **10**, 2105 (2019).

39. A. A. Albertini, C. Mérigoux, S. Libersou, K. Madiona, S. Bressanelli, S. Roche, J. Lepault, R. Melki, P. Vachette, Y. Gaudin, Characterization of monomeric intermediates during VSV glycoprotein structural transition. *PLoS Pathog.* **8**, e1002556 (2012).

40. I. S. Kim, S. Jenni, M. L. Stanifer, E. Roth, S. P. J. Whelan, A. M. van Oijen, S. C. Harrison, Mechanism of membrane fusion induced by vesicular stomatitis virus G protein. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E28–E36 (2017).

41. S. Bangaru, S. Lang, M. Schotsaert, H. A. Vanderven, X. Zhu, N. Kose, R. Bombardi, J. A. Finn, S. J. Kent, P. Gilchuk, I. Gilchuk, H. L. Turner, A. García-Sastre, S. Li, A. B. Ward, I. A. Wilson, J. E. Crowe Jr, A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface. *Cell*. **177**, 1136-1152.e18 (2019).

42. D. M. Eckert, P. S. Kim, Mechanisms of Viral Membrane Fusion and Its Inhibition. *Annu. Rev. Biochem.* **70**, 777–810 (2001).

43. J. R. Byrum, E. Waltari, O. Janson, S.-M. Guo, J. Folkesson, B. B. Chhun, J. Vinden, I. E. Ivanov, M. L. Forst, H. Li, A. G. Larson, W. Wu, C. M. Tato, K. M. McCutcheon, M. J. Peluso, T. J. Henrich, S. G. Deeks, M. Prakash, B. Greenhouse, J. E. Pak, S. B. Mehta, multiSero: open multiplex-ELISA platform for analyzing antibody responses to SARS-CoV-2 infection. *medRxiv* (2021), doi:10.1101/2021.05.07.21249238.

44. A. Glasgow, J. Glasgow, D. Limonta, P. Solomon, I. Lui, Y. Zhang, M. A. Nix, N. J. Rettko, S. Zha, R. Yamin, K. Kao, O. S. Rosenberg, J. V. Ravetch, A. P. Wiita, K. K. Leung, S. A. Lim, X. X. Zhou, T. C. Hobman, T. Kortemme, J. A. Wells, Engineered ACE2 receptor traps potently neutralize SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* (2020), doi:10.1073/pnas.2016093117.

Chapter 4

A kinetic and structural comparison of a protein's co-translational folding and refolding pathways

---

This chapter is adapted from the paper:
Samelson AJ, Bolin E, Costello SM, Sharma AK, O'Brien EP, Marqusee S. 2018. Kinetic and structural comparison of a protein's co-translational and refolding pathways. *Science Advances.* Vol 4, Issue 5.

I am the third listed author of this work. My contributions to this work include experimental and analytical work for the denaturant melt portion 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. Here, 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 the rate-limiting step of HaloTag folding is not affected by translation, 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* (*1*, *2*). In the cell, proteins are synthesized by the ribosome one amino acid at a time and the translational machinery is a major hub for protein quality control (*3*, *4*). 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 than typical refolding experiments, where the full-length protein is denatured and then allowed to refold (*5*). Cotranslational folding has thus become a highly active area of research (*6–10*), and has revealed insights into the mechanisms of protein-misfolding diseases (*11*).

Recent studies on stalled ribosome nascent chain complexes (RNCs) have illuminated some of the features that guide co-translational protein folding, specifically effects due to the tethering

and proximity of the ribosome (*12–15*). These elegant studies, however, fail to recapitulate the dynamic process of translation; isolated, stalled RNCs are not sufficient to understand the interplay between translation and protein folding. The importance of co-translational folding is highlighted by biochemical studies which have demonstrated that concurrent translation can increase the fidelity of protein folding and quaternary structure formation (*16–21*). Small changes in protein folding efficiency (the fraction of produced protein that folds to its native state) can have dramatic effects – they can overwhelm the cellular proteostasis machinery and lead to protein-misfolding diseases (*4*, *22*). Thus, proper co-translational 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. In order to understand how translation modulates protein folding, a direct comparison of the cotranslational- and re- folding trajectories is essential. However, we lack the high-resolution tools needed to monitor co-translational folding.

Here, we determine the structural basis by which co-translational 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 co-translational 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

4.3.1 HaloTag refolding can be monitored by fluorescence polarization (FP)

HaloTag is a modified haloalkane dehalogenase commonly used as a tool for *in vivo* imaging (Fig. 4-1A-B) that covalently binds ligand in its native conformation ($k = 2.7\text{x}10^6\ \text{M}^{-1}\ \text{sec}^{-1}$, ~27.0 sec^{-1} at 10μM Tetramethylrhodamine-ligand (TMR), the concentration used in this study) (*23*). If folding is much slower than 27.0 sec^{-1}, 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.

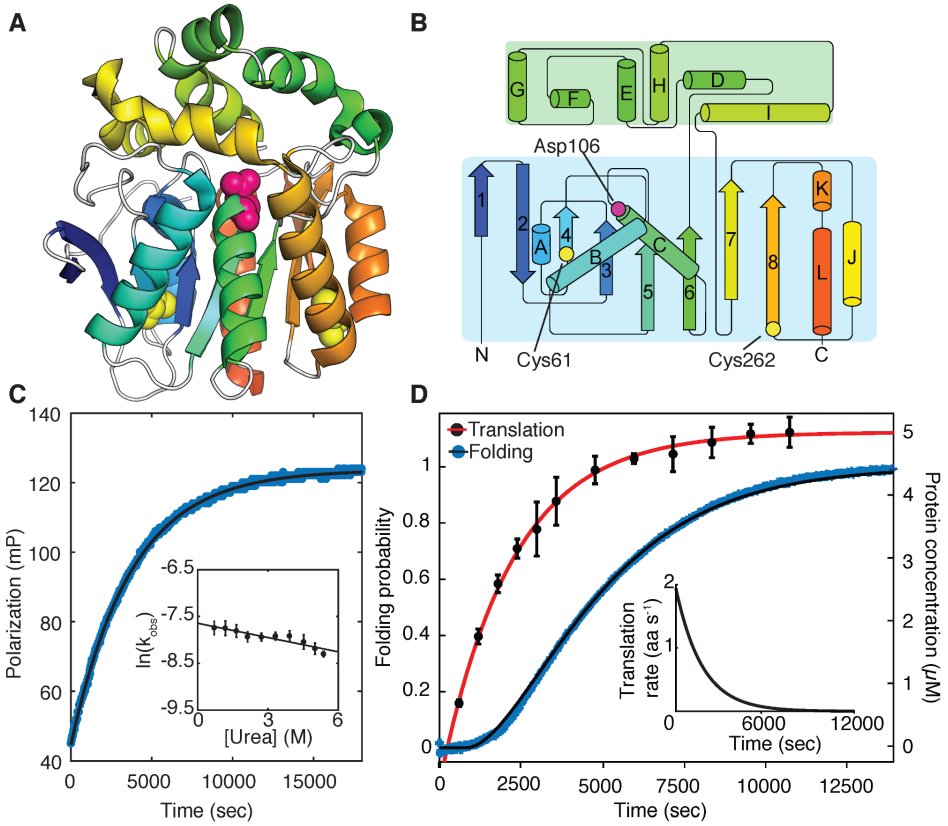


Figure 4-1. **Cotranslational folding of HaloTag monitored in real-time with fluorescence polarization. (A)** Cartoon representation of the crystal structure of HaloTag from PDB 5UY1 (*39*). Cysteines 61 and 262, as well as the active-site Asp (D106), are represented as spheres **(B)** Secondary structure map of HaloTag (*40*) **(C)** Refolding of HaloTag as a function of time monitored by fluorescence polarization. HaloTag refolding as a function of denaturant concentration (inset). **(D)** HaloTag cotranslational folding (blue; left axis) measured by FP, and HaloTag synthesis (black circles with red fit; right axis), measured by gel (see Figure 4-2). Elongation rate as a function of time is shown inset.

We monitored refolding of HaloTag in the presence of 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 (e.g. 8M urea to 0.8M urea), results in single exponential kinetics (k_{obs} = 4.7±0.9x10^{-4} sec^{-1} at a final urea concentration of 0.8M urea, Fig. 4-1). Linear extrapolation of the natural log of the folding rate, ln(k_{obs}), as a function of the final urea concentration (*24*) yields a folding rate in the absence of denaturant, $k_{f,H2O,FP}$ = 4.8±0.6x10^{-4}sec^{-1}, >10,000 times slower than ligand binding (Table 4-1) and similar to the folding rate determined by circular dichroism (see below). Thus, changes in FP are a measure of HaloTag folding.

Thermodynamic Data		
	ΔG_{unf} (kcal · mol^{-1}) (CD)	6.03±0.39
	m-value (kcal · mol^{-1}M^{-1})	1.57±0.11
Kinetic Data		
	ΔG_{unf} (kcal · mol^{-1})	5.24±2.0
	m-value (kcal · mol^{-1}M^{-1})	1.41±0.58
	$k_{f,H2O}$ (sec^{-1})	0.04±0.02
	m_f (kcal · mol^{-1}M^{-1})	1.46±0.71
	$k_{constant, H2O}$ (sec^{-1})	6.6±0.71x10^{-4}
	$m_{constant}$ (kcal · mol^{-1}M^{-1})	0.02±0.1
	$k_{f, H2O, FP}$ (sec-1)	4.8±0.6x10^{-4}
	m_{FP} (kcal · mol^{-1}M^{-1})	-0.1±0.04
	$k_{NI, H2O}$ (sec^{-1})	8.47±20±10^{-6}
	m_{NI} (kcal · mol^{-1}M^{-1})	-0.44±0.3
	$k_{IU, H2O}$ (sec^{-1})	3.3±9.9x10^{-4}
	m_{IU} (kcal · mol^{-1}M^{-1})	0.70±.15
Cysteine Accessibility		
	$k_{WT,refolding, slow}$ (sec^{-1})	7.8±0.6±10^{-4}
	$k_{WT,refolding, fast}$ (sec^{-1})	0.03±.02
	$k_{WT,IVT}$ (sec^{-1})	4.7±0.3x10^{-4}
	$k_{M129C,refolding}$ (sec^{-1})	>0.01
	$k_{M129C,IVT}$ (sec^{-1})	3.2±0.2x10^{-4}
	$k_{I126C,refolding}$ (sec^{-1})	>0.01
	$k_{I126C,IVT}$ (sec^{-1})	2.2±0.2x10^{-4}
	$k_{E121C,refolding}$ (sec^{-1})	>0.01
	$k_{E121C,IVT}$ (sec^{-1})	>0.01

Table 4-1: Summary of kinetic and thermodynamic data of HaloTag and its mutants (error bars are standard deviations of at least three separate measurements).

4.3.2 Co-translational folding can be monitored by fluorescence polarization

In order to monitor co-translational 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 *in vitro* transcription/translation system (NEB PURExpress). Figure 4-1D reveals biphasic kinetics: a lag phase and an exponential phase. Importantly, the observed kinetics are independent of the TMR-ligand concentration and are specific for the HaloTag gene (Fig. 4-2, Table 4-2). To confirm that the changes in TMR polarization monitor co-translational folding and not protein synthesis, we determined the time-dependence of protein production independently using a gel-based assay (Fig. 4-1D) (*25*). 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 fluorescence polarization reports on HaloTag folding and is not translation-limited (Fig. 4-2).

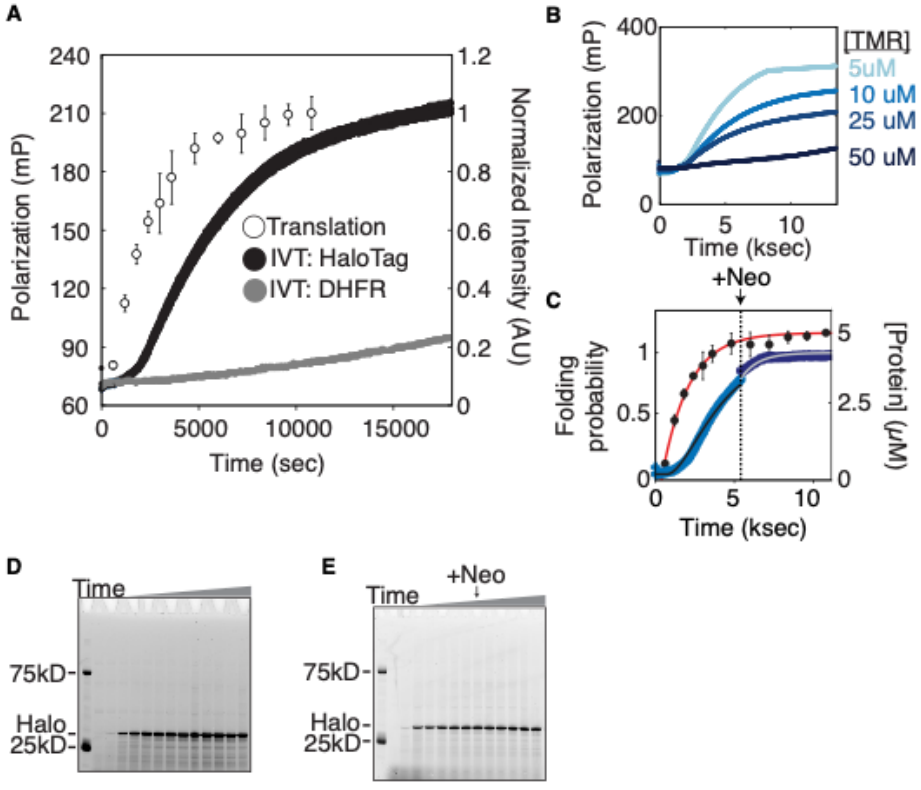


Figure 4-2. Cotranslational folding of HaloTag can be measured using fluorescence polarization. (A) Raw FP data (left axis) for IVT reactions initiated with HaloTag (black) and DHFR (gray) plasmids. Translation (right axis) of HaloTag as determined by gel. **(B)** Polarization as a function of TMR-ligand 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 (A) and (C).

	Lag Time (sec)	Rate ($x10^{-4}$ sec^{-1})
Refolding (polarization)	NA	4.8±0.6
Folding (in vitro translation)	811.29±8.57	4.42±0.02
Translation	251.29±34.6	4.50±0.10
$Folding_{before +neo}$	724.18±17.0	5.49±0.07
$Folding_{after +neo}$	NA	9.13±0.19
$Folding_{[TMR] = 5uM}$	1065.89±17.32	4.39±0.04
$Folding_{[TMR] = 10uM}$	811.29±8.6	4.42±0.02
$Folding_{[TMR] = 12.5uM}$	780.05±8.9	6.13±0.04

Table 4-2. Kinetic data obtained for HaloTag folding using fluorescence polarization.

4.3.3 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 methods). The protein synthesis data were analyzed to determine the translation lag phase (251±35 seconds), which represents the time to synthesize detectable protein levels, and the time-dependent translation rate**.** The average translation rate, ~1aa/sec, is similar to translation rates determined for other *in vitro* systems (*26*, *27*). To determine the co-translational folding rate, we augmented this model to include a protein-folding component. The lag time observed for the change in polarization is four-fold larger than that observed for protein synthesis, 811±9 versus 251±35 seconds, respectively. The resulting co-translational folding rate is thus $4.42±0.02x10^{-4}sec^{-1}$, similar to that obtained in the absence of translation (Table 4-1).

4.3.4 Refolding studies of HaloTag

We then characterized the stability and refolding of HaloTag using recombinant, purified protein (Fig. 4-3; Table 4-1). Folding kinetics, determined by CD, fit to two exponential phases, a fast-folding phase and a urea-independent slow phase (Fig. 4-3A-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 cis-trans proline isomerization. However, both refolding and co-translational folding in the presence of the proline isomerase cyclophilin A revealed no effect, which suggests this may not be due to proline isomerization (Fig. 4-4, Tables 4-1 and 4-3). Surprisingly, refolding to below 1.0M urea resulted in visible precipitation and protein aggregation (Fig. 4-3 and 4-5) even though no aggregation was observed in the above

co-translational folding experiments that take place at 0M urea. Aggregation occurred after an initial decrease in CD signal with a rate similar to the fast refolding phase observed in non-aggregating conditions. Using centrifugation, we determined the fraction of soluble protein to be 0.70±0.06 under these conditions (Fig. 4-6).

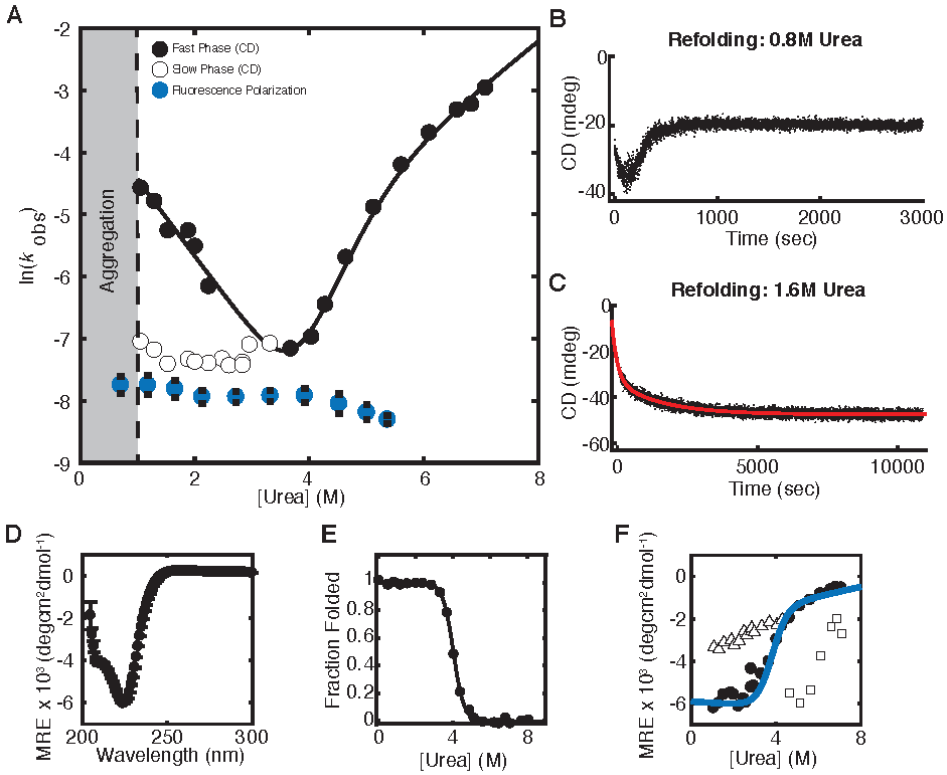


Figure 4-3. Characterization of HaloTag folding kinetics and stability. (A) Chevron plot of HaloTag folding and unfolding rates as a function of urea concentration. Fast phase (black circles) and slow phase (white circles, black outline). Refolding as measured by FP is shown in blue. Refolding traces of HaloTag at 0.8M urea, where there is visible protein aggregation **(B)**, and 1.6M urea, **(C),** where no precipitation is observed. **(D)** CD spectrum of HaloTag at 0M 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.

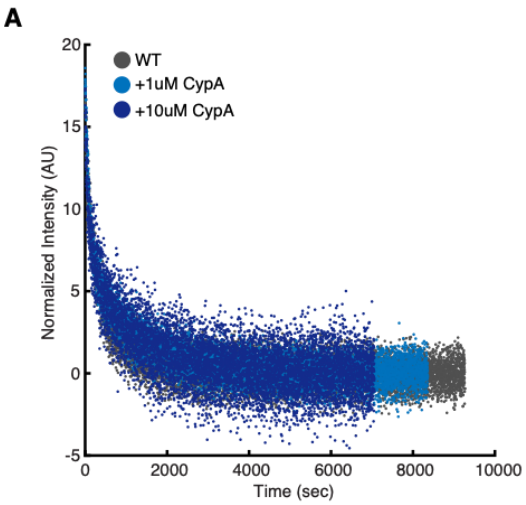
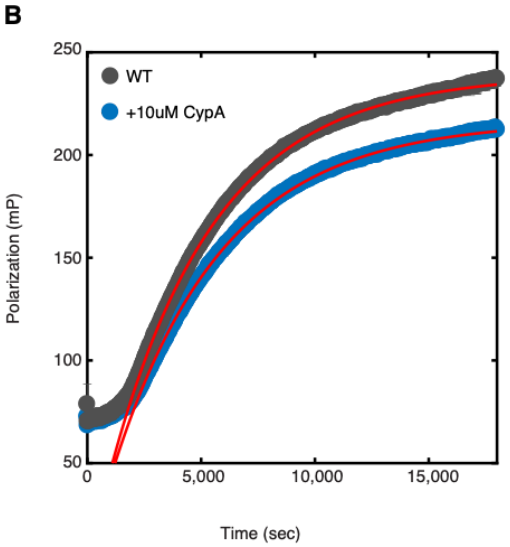


Figure 4-4. Addition of the peptidyl-proline isomerase CypA does not affect HaloTag refolding or co-translational folding rates. (A) Refolding of HaloTag in increasing concentrations of CypA as monitored by CD. **(B)** FP of HaloTag in the presence of 10mM CypA (blue) and no CypA (gray).



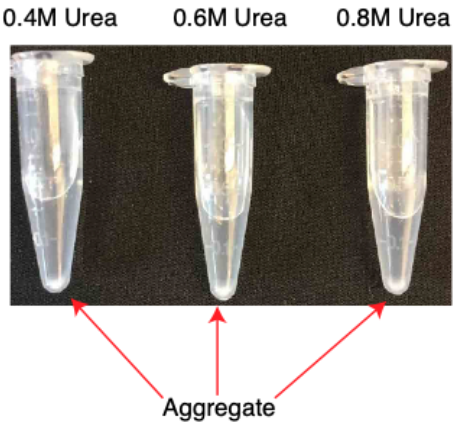


Figure 4-5. Aggregation of HaloTag. HaloTag aggregates after refolding via dilution from 8.0M urea to the indicated final concentrations of urea.

	Rate ($x10^{-4}$ sec^{-1})	t_0 (sec)
WT (10μM TMR)	4.40±0.02	811.3±9
+10μM CypA	2.12±0.02	2007±70
Halo** I126C	2.08±0.02	612±33
Halo** M129C	2.20±0.08	848±10
Halo** E121C	2.32±0.11	755±57

Table 4-3. Folding rates of HaloTag and variants measured by FP.

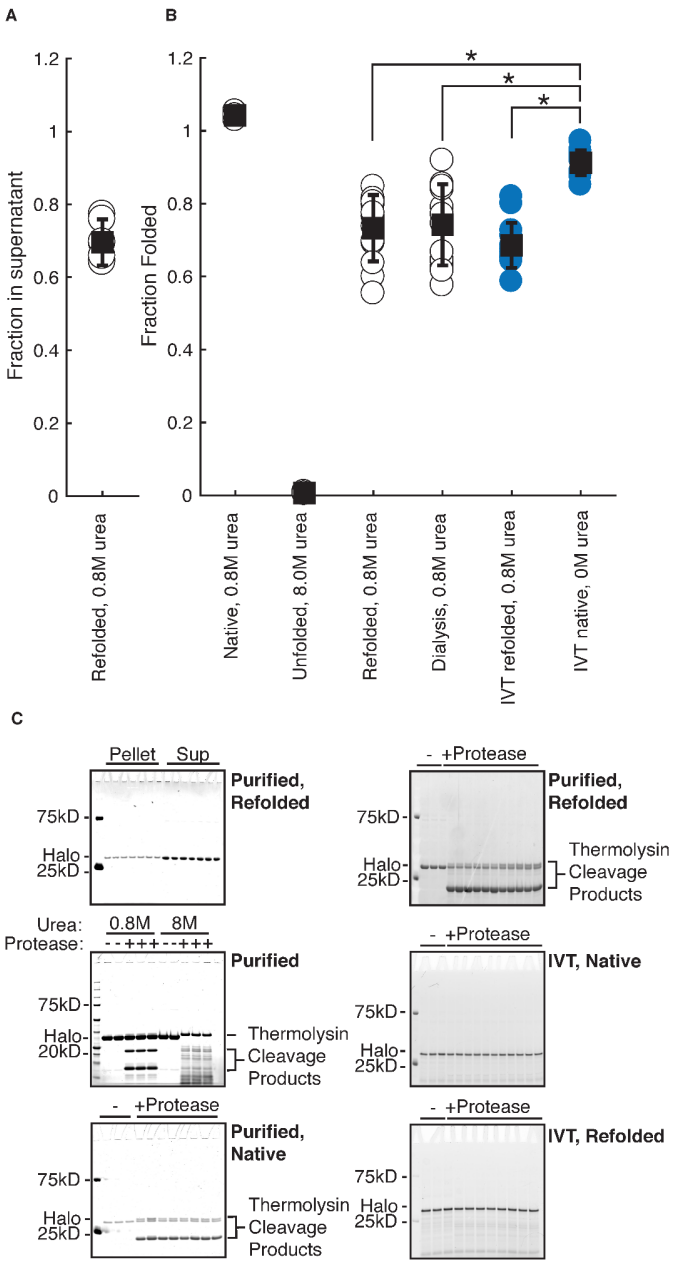


Figure 4-6. HaloTag folding is more efficient during *in vitro* translation than after refolding. **(A)** Fraction of total protein remaining in supernatant after centrifugation following refolding of HaloTag to 0.8M urea. **(B)** Fraction folded as measured by pulse proteolysis in conditions as indicated; either after refolding, after *in vitro* translation, or both. Blue circles are *in vitro* translated protein. **(C)** Representative gels for panels **(A)** and **(B)**. All error bars are the SD of at least 15 separate experiments except for HaloTag in 0.8 and 8.0M urea, which are the SD of three experiments. **'*'** represents a p-value of less than 0.01, using a student's unpaired t-test.

4.3.5 HaloTag co-translational folding is more efficient than refolding

To compare the efficiencies of refolding and co-translational folding, i.e. the fraction of protein that reaches the native state, we used pulse proteolysis (*28*), a gel-based method for measuring the amount of folded protein (Fig. 4-6). Purified HaloTag incubated in 0.8 M urea is completely folded when evaluated by pulse proteolysis. But, when refolded by dilution from 8.0M to 0.8M urea, the efficiency is only 0.73±0.03, consistent with that determined by centrifugation above. In contrast, co-translational folding is significantly more efficient than refolding: 0.91±0.03 versus 0.73±0.03 ($p<.01$, student's unpaired t-test; $n > 12$; Fig. 4-6, Table 4-4). It is important to note that IVT reactions are carried out at a higher protein concentration than the less-efficient refolding studies (>5uM and 3uM, respectively; see Fig. 4-2 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. 4-6). Why is co-translational folding significantly more efficient than refolding; how does translation alter the folding pathway of HaloTag?

	Fraction Folded	Number of samples
Native	1.04±0.01	3
Unfolded	0.0006±0.0003	3
0.8M refolded	0.73±0.10	15
0.8M dialysis	0.74±0.11	15
IVT refolded	0.69±0.06	15
IVT native	0.91±0.03	15

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

4.3.6 Structural characterization of the in vitro refolding pathway using HXMS

To compare the refolding and co-translational 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 conformations formed during HaloTag refolding (*29*, *30*). Pulses of hydrogen exchange are applied at various refolding times and the mass of individual peptides is monitored as a function of refolding time. Changes in mass are a measure of backbone amide accessibility at that particular refolding time. Figure 4-7A shows the fraction deuterated for each peptide after ten seconds of refolding and reveals two populations: those that are at least 25% deuterated by ten seconds ("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. 4-7B). On average, "fast" peptides are more deuterated at all time points analyzed than "slow" peptides. These data, together with the biphasic CD kinetics (Fig. 4-3), 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, as well as β-strand 8, are protected more slowly (Fig. 4-7C). 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, 2 and Helix C) (Fig, 4-7C). 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, since this intermediate is comprised of residues distant in sequence space, it is possible that co-translational folding does not involve formation of this intermediate.

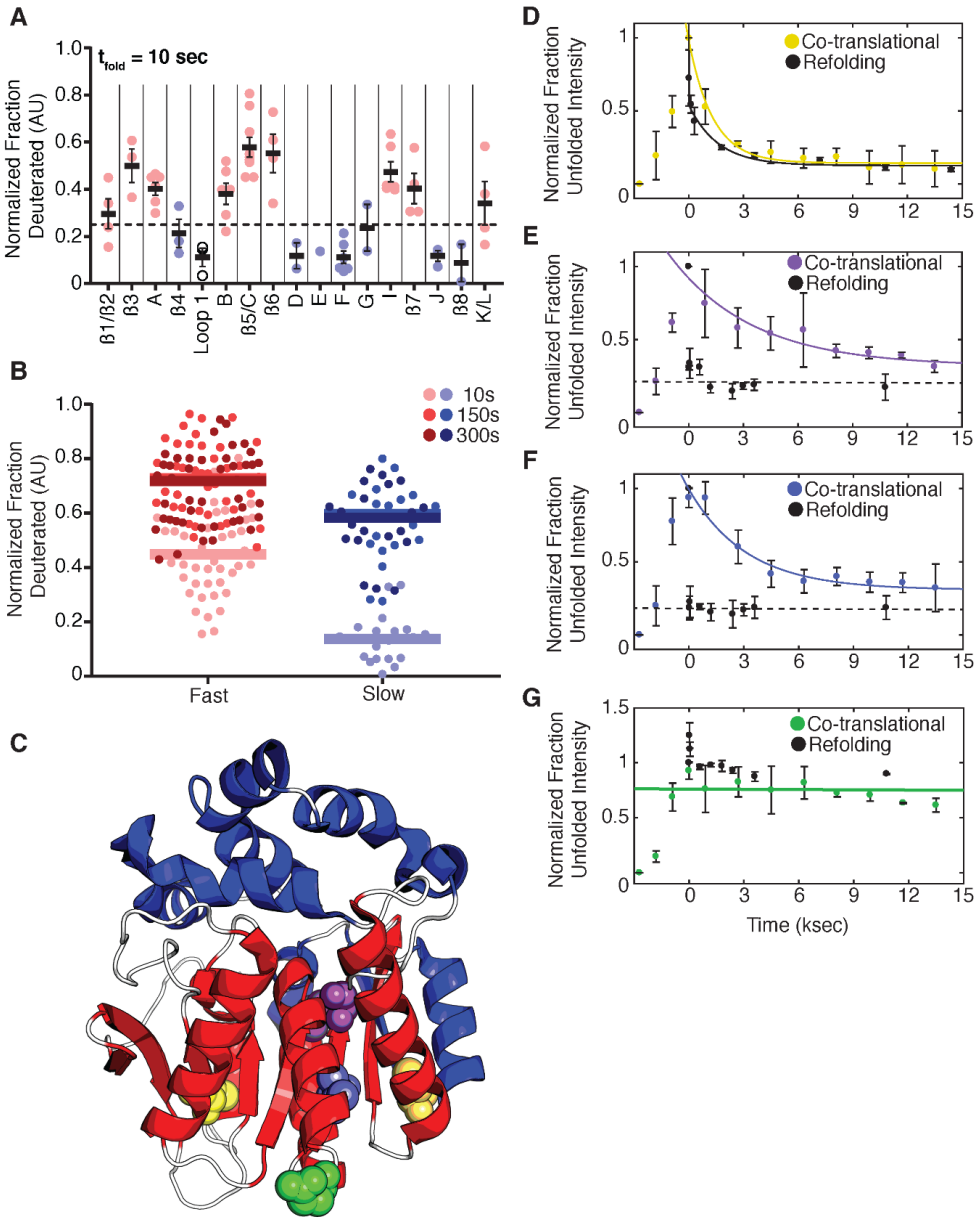


Figure 4-7. The HaloTag folding trajectory changes during cotranslational folding. (A) Peptides derived from HX-MS experiments after ten seconds of refolding were plotted according to their corresponding secondary structural element. Helices are lettered while beta-sheets are numbered. Secondary structural elements were then divided into fast (red circles) or slow (blue circles) folding regions based on the average fraction deuterated (solid line) for peptides within those secondary structures at the ten-second time point (above or below dashed line). Error bars represent the 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. **(C)** Crystal structure of HaloTag with slow (blue) and fast (red) secondary structural elements colored. Loops are colored in white. Cysteines probed in D-G are represented as spheres: Yellow – Cys61 and Cys262; Purple – M129C; Blue – I126C; Green – E121C. **(D)-(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.

4.3.7 Comparison of the HaloTag refolding and co-translational folding trajectories using pulsed cysteine labeling

Unfortunately, the above HX-MS studies are currently not feasible for investigating co-translational folding. Therefore, to determine if HaloTag populates the same folding intermediate during co-translational 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 (*7*). We monitored thiol accessibility during both refolding and cotranslational folding using a fluorescein-conjugated maleimide (F5M), detected by in-gel fluorescence.

Both of HaloTag's native cysteines, positioned at the base of two beta-strands, β4 and β8, are completely protected in the folded state and accessible in the unfolded state (modified within 30 seconds; Fig. 4-8). 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 protein showed two phases with similar rates to those obtained by CD. In contrast, pulsed-thiol labeling during co-translational folding resulted in only a single exponential indistinguishable from HaloTag's slow folding rate during refolding (Fig. 4-7D, Table 4-1).We then created three site-specific cysteine variants to probe the very early stages of folding (E121C, I126C, 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. 4-7E and F). E121C is on the surface of HaloTag and remains unprotected throughout the folding reaction (Fig. 4-7G). All three variants bind TMR and display similar folding kinetics as WT HaloTag (Fig. 4-9). In contrast to the previous refolding experiments, during *in vitro* translation, sites 126 and 129 are not protected early, but rather show slow protection corresponding to the overall folding rate of the protein (Fig, 4-7D). Thus, the folding pathway of HaloTag is altered during translation.

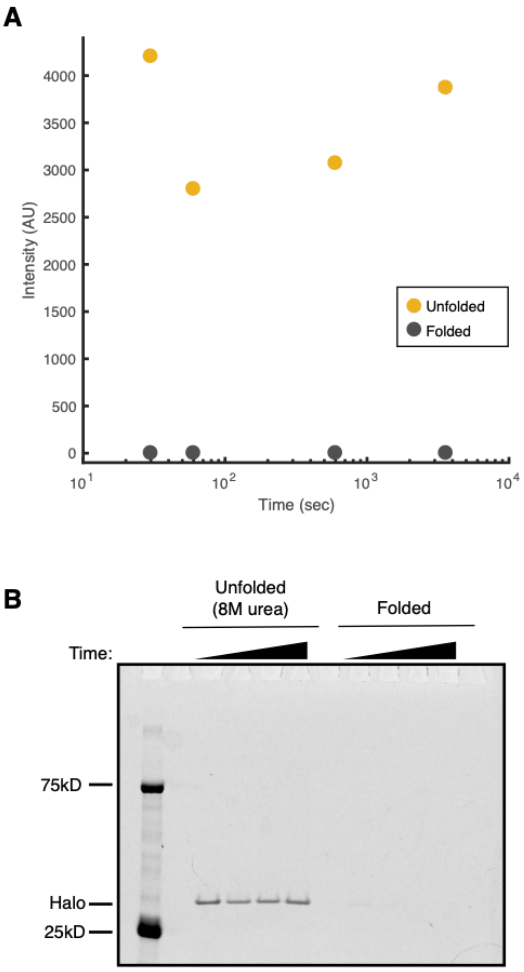


Figure 4-8. Cysteine accessibility of WT HaloTag. (A) Cysteine accessibility as a function of time as measured by fluorescein-maleimide fluorescence for unfolded (yellow circles) and folded (grey circles) HaloTag. **(B)** Raw data for plot in (A)

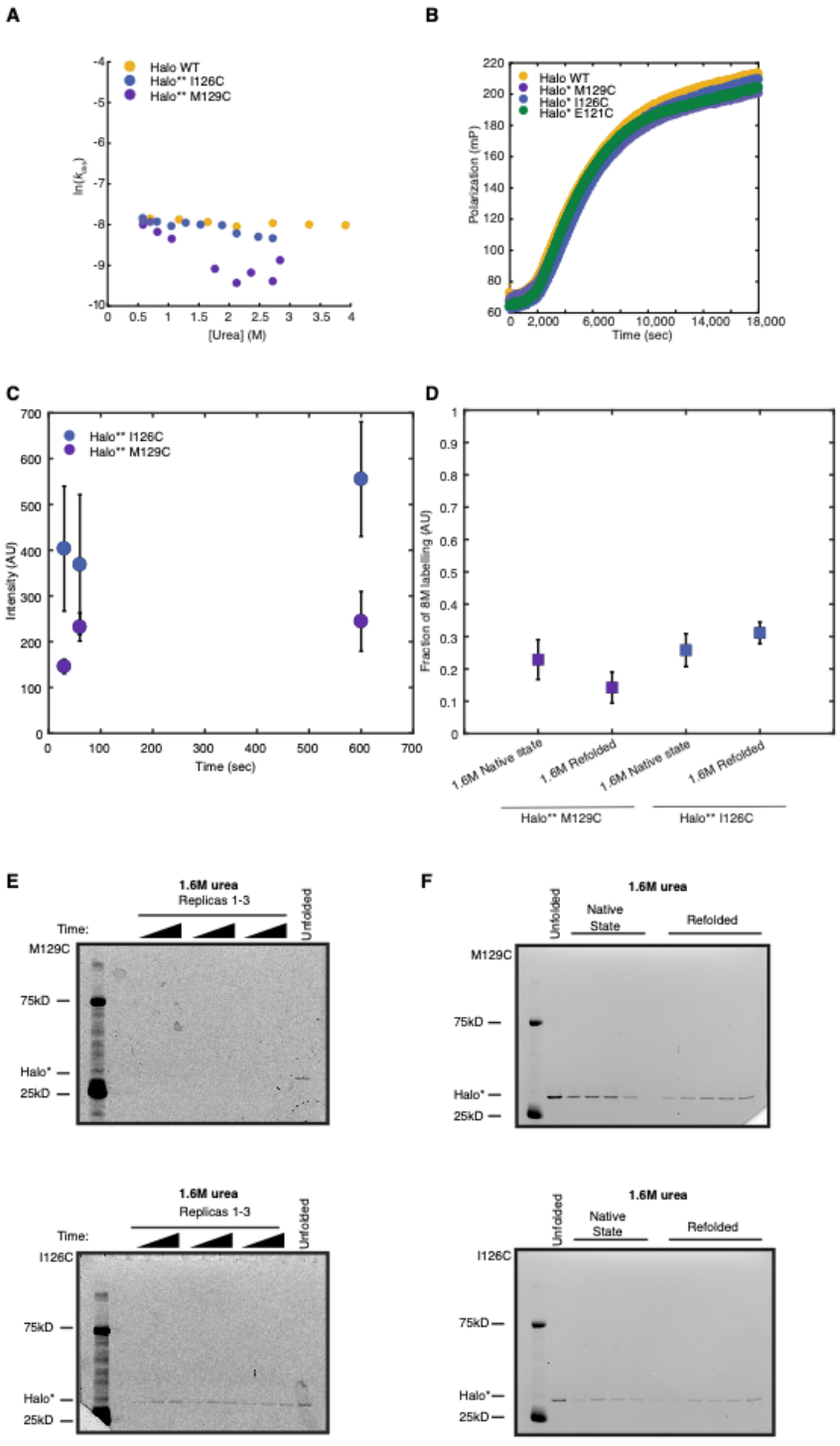


Figure 4-9. Characterization of Halo* cysteine mutants. (A) Refolding rate as a function of [urea] as measured by FP for different HaloTag constructs. WT – yellow; Halo* M129C – purple; Halo* I126C – blue. **(B)** Cotranslational folding of HaloTag variants measure by FP. WT – yellow; Halo* M129C – purple; 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

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 co-translational 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 (*31*). The specific cysteines characterized here, however, do not yield further insight into other potential intermediates that may form during translation. Interestingly, the overall rate of folding is not changed during co-translational 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 (*32*, *33*). 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 (*12*, *34*, *35*). 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 (*36–38*). 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 Methods

4.5.1 Protein expression and purification

BL21(DE3) cells were transformed with expression vectors containing the wild-type or mutant HaloTag cDNA. Single colonies were used to seed starter cultures grown overnight to saturation. Large-scale cultures were inoculated with 5mL of overnight culture, grown at 37°C to an OD600 of 0.6-0.8 and induced with 1 mM IPTG for 2-3 hours at 37°C. After induction, cultures were pelleted at 5000 x g for 10 minutes at 4°C, flash frozen, and stored at -80°C.

Cell pellets were resuspended in 10 mM Tris/H_2SO_4, pH 7.5, 1 mM TCEP (Lysis Buffer) and lysed by sonication on ice. Lysates were cleared by centrifugation for 30 minutes at 20,000 xg, 4°C and subsequently filtered through 0.2 µm filters. After clearing, lysate was dialyzed into at least a 10-fold volume excess of Lysis Buffer, loaded onto a HiPrep 16/10 Q Xl column equilibrated with Lysis Buffer and eluted with a gradient of Lysis Buffer plus 0 to 600mM NaCl. Fractions containing the HaloTag protein were dialyzed into at least a 10-fold volume excess of 20mM Sodium Acetate, pH 5.0 (Q Buffer) loaded onto a HiPrep 16/10 Q Xl column equilibrated with Q Buffer and eluted with a gradient of Q Buffer plus 0 to 800mM NaCl. Fractions containing HaloTag protein were then concentrated and purified on a HiLoad 16/600 Superdex 75 pg column equilibrated with 50mM ammonium bicarbonate or 25mM HEPES KOH pH 7.5, 15mM MgOAc, 150mM KCl, 0.1mM 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.

4.5.2 Fluorescence polarization

Data collection: All experiments were performed at 37°C unless otherwise noted. Fluorescence polarization 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 as well as a 530nm/590nm filters with side gain set at 45 and top gain set at 40. Read height was 7.5mm and 10 measurements were made per data point. After a 15 minute incubation at 37°C, readings were initialized by the addition of DNA (IVT reactions) or unfolded protein (refolding experiments). Measurements were taken every 20 seconds for five hours after thirty seconds of mixing and a 90 second delay for temperature equilibration.

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

Refolding experiments: Refolding experiments were performed in HKMT buffer plus appropriate concentrations of urea and TMR (to a final concentration of 5μM TMR, 3.33% DMSO). Plates were sealed and incubated at 37°C for fifteen minutes until reactions were initiated by adding 10μL of 20μM HaloTag in 8M 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 (*28*).

4.5.3 Circular dichroism

Kinetic and equilibrium experiments were performed using a 0.5 cm cuvette at 37°C with constant stirring at 3μM (0.1mg/mL) in HKMT buffer. Equilibrium and kinetic experiments were performed as previously described (*41*), but at a wavelength of 225nm instead of 222nm to increase signal-to-noise. Analysis was performed as described (*41*).

Wavelength experiments were performed in a 0.1cm cuvette at 37°C with 15μM protein (~0.5mg/mL) in HKMT buffer.

4.5.4 Determination of folding efficiency

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

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

Pulse Proteolysis: IVT reactions were performed as per the manufacturer's instructions but with the addition of 1μL of RNase inhibitor, murine and 1.25μL Fluorotect Greenlys (Promega) per 30μL IVT reaction (*25*). IVT reactions were quenched after 1 hour to a final concentration of 2mM Chloramphenicol and 0.1mg/mL RNase A. 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 10μL and 1μL of 1mg/mL Thermolysin (Sigma) was added to each reaction for 1 minute and quenched with EDTA to a final concentration of 83mM. SDS-PAGE loading dye was then added to each reaction and each reaction run on a 4-12% Bis Tris gel in MES run buffer. Imaging and analysis was performed as described previously (*28*).

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

4.5.5 Translation rate measurement

IVT reactions were performed as per the manufacturer's instructions but with the addition of 1µL of RNase inhibitor, murine and 1.25µL Fluorotect Greenlys (Promega) per 30µL IVT reaction and initiated with 250ng/25µL IVT reaction of DNA. At each time point, 1.5µL of IVT reaction were quenched into a final concentration of 2mM Chloramphenicol and 0.1mg/mL RNase A and then SDS-PAGE loading dye. Reactions were then run on a 4-12% Bis Tris gel in MES run buffer and imaged using a Typhoon Trio. Analysis was performed using ImageJ.

4.5.6 Cysteine protection assays

Purified protein: Refolding reactions were initiated as described above. At each time point, a 50-fold molar excess of fluorescein-maleimide (F5M) was added for 30 seconds and quenched into an equal volume of SDS-PAGE loading dye containing Beta-mercaptoethanol to a final BME concentration of 2.15M. Reactions were then run on 4-12% Bis Tris gel in 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):

$$f(t) = a*(1-e^{-kt})+c$$

Since 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 Figure 4-10.

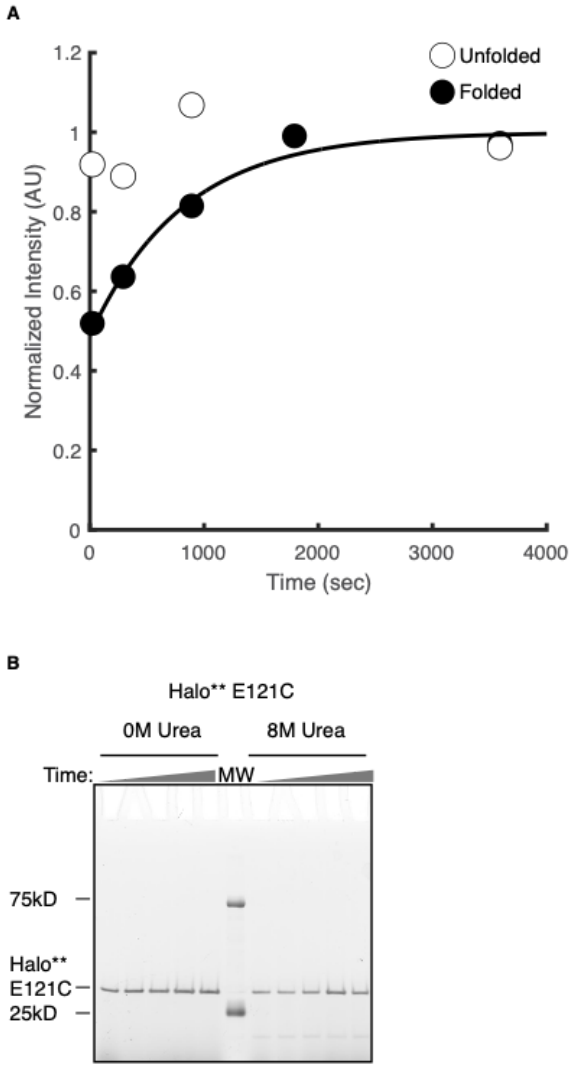


Figure 4-10. Characterization of Halo* E121C cysteine accessibility (A) Cysteine accessibility of folded (black circles) and unfolded (white circles) Halo* E121C as measured by fluorescein-maleimide reactivity. **(B)** Gel used in (A)

IVT reactions: IVT reactions were initiated as described above. At each time point, an equal volume of 2mM F5M was mixed with IVT reaction for 30 seconds and quenched into SDS-PAGE loading dye as above. At 45 minutes, reactions were halted by the addition of chloramphenicol to a final concentration of 2mM. Reactions were then run on a 4-12% Bis Tris gel in MES run buffer and imaged using a Typhoon Trio. A sample of purified, TMR-bound HaloTag was run to determine the size of the Halo-F5M band. Analysis was performed using ImageJ. Intensities were normalized to a major protein product running at ~65kD 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 minutes, using bi-square fitting and "*k*" bounded at zero:

$$f(t) = a*(1-e^{-kt})+c$$

4.5.7 Pulsed labeling HX-MS

Pulse-labeling scheme: The HX-MS pulsed labeling experiments were based on previously described approaches (*29*, *30*). Deuterated protein was prepared by lyophilizing unfolded HaloTag in 8M urea followed by resuspension in D_2O, repeated four times. Refolding experiments were carried out using a Bio-logic QFM-4 apparatus in interrupted flow mode. To initiate refolding, deuterated protein in 8M urea (deuterated HKMT buffer, pD_{read} 7.9, 10°C) was diluted with 10 volumes of deuterated HKMT to a final urea concentration of 1.6M. 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, 10ms, 5 volumes; final pH=10.00). Due to the large volume changes, the final solution is only 31% protonated. The pulse was quenched by dilution with a low pH buffer (1M Glycine, 5 volumes; final pH=2.00) to slow any further exchange. Protein samples were then collected and injected into a custom 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 HPLC 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 Proteome Discoverer 2.0 software. Peptide mass envelopes were fit using HDExaminer (Sierra Analytics) followed by manual confirmation of each peptide. Deuterium content was assessed by examination of the centroid of each fitted peptide mass envelope. Only peptides with high signal/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.

4.5.8 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

$$L \rightarrow P. \qquad [1]$$

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

$$B \rightarrow P, \quad [2]$$

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

$$\frac{d[P(t)]}{dt} = k[B(0) - P(t)]. \quad [3]$$

Solving Eq. [3] under boundary conditions, $[P(t)] = 0$ at $t = 0$, yields

$$[P(t)] = [B(0)](1 - e^{-kt}). \quad [4]$$

The maximum protein concentration $[P_{max}]$ at time $t \rightarrow \infty$ is equal to $[B(0)]$. Therefore,

$$[P(t)] = [P_{max}](1 - e^{-kt}). \quad [5]$$

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

$$[P(t)] = [P_{max}](1 - e^{-k(t-t_0)}). \quad [6]$$

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 (*42*). We calculated the rate of protein production $J(t)$ by taking the derivative of Eq. [6]

$$J(t) = [P_{max}]ke^{-k(t-t_0)} \quad [7]$$

Next, we normalized this quantity with the ribosome concentration (*i.e.*, 0.45 μM), 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.

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

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

and

$$[U(t)] + [F(t)] = [P(t)]. \quad [10]$$

$[F(t)]$ 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

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

and

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

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

4.6 Acknowledgments

We would like to thank the entire Marqusee lab for advice on experiments, paper writing and support. Especially S.A. Lim for help with kinetic fitting and M.K. Jensen for help with the manuscript. Also, we would like to thank A. Martin and his lab for equipment and reagents and especially D. Costanzo for advice relating to F5M labeling. J.S. Fraser for advice on the manuscript. J. Chen for technical support. This work was supported by NIH grant GM050945 (AJS, EB, SMC, SM) NIH grant 428-21 59ZN and NSF grant 428-21 60BK (AK, EPO).

4.7 References

1. E. Braselmann, J. L. Chaney, P. L. Clark, Folding the proteome. *Trends Biochem. Sci.* **38**, 337–44 (2013).

2. F. U. Hartl, M. Hayer-Hartl, Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574–81 (2009).

3. S. Pechmann, F. Willmund, J. Frydman, The ribosome as a hub for protein quality control. *Mol. Cell*. **49**, 411–21 (2013).

4. D. Balchin, M. Hayer-Hartl, F. U. Hartl, In vivo aspects of protein folding and quality control. *Science* **353 (**2016).

5. E. P. O'Brien, J. Christodoulou, M. Vendruscolo, C. M. Dobson, New scenarios of protein folding can occur on the ribosome. *J. Am. Chem. Soc.* **133**, 513–26 (2011).

6. L. D. Cabrita, C. M. Dobson, J. Christodoulou, Protein folding on the ribosome. *Curr. Opin. Struct. Biol.* **20**, 33–45 (2010).

7. J. Lu, C. Deutsch, Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **12**, 1123–9 (2005).

8. G. Kramer, D. Boehringer, N. Ban, B. Bukau, The ribosome as a platform for cotranslational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **16**, 589–97 (2009).

9. O. B. Nilsson *et al.*, Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. & Mol. Biol.* **24**, 221–226 (2017).

10. O. B. Nilsson *et al.*, Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep.* **12**, 1533–1540 (2015).

11. S. J. Kim *et al.*, Translational tuning optimizes nascent protein folding in cells. *Science*. **348**, 444–8 (2015).

12. A. J. Samelson, M. K. Jensen, R. A. Soto, J. H. D. Cate, S. Marqusee, Quantitative determination of ribosome nascent chain stability. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13402–13407 (2016).

13. A. M. Knight *et al.*, Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem. Biol.* **8**, 1195–204 (2013).

14. C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco, C. Bustamante, The ribosome modulates nascent protein folding. *Science*. **334**, 1723–7 (2011).

15. L. D. Cabrita *et al.*, A structural ensemble of a ribosome-nascent chain complex during

cotranslational protein folding. *Nat. Struct. Mol. Biol.* (2016), doi:10.1038/nsmb.3182.

16. J. Labbadia, R. I. Morimoto, The Biology of Proteostasis in Aging and Disease. *Annu. Rev. Biochem.*, 1–30 (2015).

17. G. Zhang, M. Hubalewska, Z. Ignatova, Transient ribosomal attenuation coordinates protein synthesis and cotranslational folding. *Nat. Struct. Mol. Biol.* **16**, 274–80 (2009).

18. K. G. Ugrinov, P. L. Clark, Cotranslational folding increases GFP folding yield. *Biophys. J.* **98**, 1312–20 (2010).

19. J. Frydman, H. Erdjument-Bromage, P. Tempst, F. U. Hartl, Cotranslational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat. Struct. Biol.* **6**, 697–705 (1999).

20. A. N. Fedorov, T. O. Baldwin, Process of biosynthetic protein folding determines the rapid formation of native structure. *J. Mol. Biol.* **294**, 579–586 (1999).

21. Y.-W. Shieh *et al.*, Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science*. **350**, 678–680 (2015).

22. F. Gloge, A. H. Becker, G. Kramer, B. Bukau, Cotranslational mechanisms of protein maturation. *Curr. Opin. Struct. Biol.* **24**, 24–33 (2014).

23. G. V Los *et al.*, HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–82 (2008).

24. A. Matouschek, J. T. Kellis, L. Serrano, A. R. Fersht, Mapping the transition state and pathway of protein folding by protein engineering. *Nature*. **340**, 122–6 (1989).

25. A. L. Mallam, S. E. Jackson, Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat. Chem. Biol.* **8**, 147–53 (2012).

26. C. C. Guet *et al.*, Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic Acids Res.* **36** (2008), doi:10.1093/nar/gkn329.

27. G. Zubay, In vitro synthesis of protein in microbial systems. *Annu. Rev. Genet.* **7**, 267–87 (1973).

28. C. Park, S. Marqusee, Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat. Methods*. **2**, 207–12 (2005).

29. W. Hu *et al.*, Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7684–9 (2013).

30. L. Mayne *et al.*, Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation-mass spectrometry method. *J. Am. Soc. Mass Spectrom.* **22**, 1898–1905 (2011).

31. Y. Liu *et al.*, AgHalo: A Facile Fluorogenic Sensor to Detect Drug Induced Proteome Stress. *Angew. Chem. Int. Ed. Engl.* (2017), doi:10.1002/anie.201702417.

32. T. R. Jahn, S. E. Radford, Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**, 100–117 (2008).

33. K. H. Lim *et al.*, Structural Changes Associated with Transthyretin Misfolding and Amyloid Formation Revealed by Solution and Solid-State NMR. *Biochemistry*. **55**, 1941–1944 (2016).

34. J. D. Hulleman, W. E. Balch, J. W. Kelly, Translational attenuation differentially alters the fate of disease-associated fibulin proteins. *FASEB J.* **26**, 4548–60 (2012).

35. F. Trovato, E. P. O'Brien, Fast Protein Translation Can Promote Co- and Posttranslational Folding of Misfolding-Prone Proteins. *Biophys. J.* **112**, 1807–1819 (2017).

36. G. Zhang, Z. Ignatova, Folding at the birth of the nascent chain: Coordinating translation with cotranslational folding. *Curr. Opin. Struct. Biol.* **21** (2011), pp. 25–31.

37. S. Pechmann, J. Frydman, Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* **20**, 237–43 (2012).

38. M. S. Evans, I. M. Sander, P. L. Clark, Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J. Mol. Biol.* **383**, 683–92 (2008).

39. Y. Liu *et al.*, The Cation-π Interaction Enables a Halo-Tag Fluorogenic Probe for Fast No-Wash Live Cell Imaging and Gel-Free Protein Quantification. *Biochemistry*. **56**, 1585–1595 (2017).

40. A. Stivala, M. Wybrow, A. Wirth, J. C. Whisstock, P. J. Stuckey, Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics*. **27**, 3315–6 (2011).

41. K. B. Connell, E. J. Miller, S. Marqusee, The folding trajectory of RNase H is dominated by its topology and not local stability: a protein engineering study of variants that fold via two-state and three-state mechanisms. *J. Mol. Biol.* **391**, 450–60 (2009).

42. M. C. Capece, G. L. Kornberg, A. Petrov, J. D. Puglisi, A simple real-time assay for in vitro translation. *RNA*, 1–10 (2014).

Chapter 5

Future directions for the biophysical and structural investigation of nascent chains during translation

This chapter serves as a summary and future outlook for the co-translational protein folding field. The work presented in this chapter was led by me. Single-molecule fluorescence experiments were performed in close collaboration with Arjun Prabhakar and Joseph D. Puglisi at Stanford University.

5.1 Introduction.

As discussed in Chapter 1, a central goal of the work presented in this dissertation is to develop tools and approaches that will allow us to ask detailed biophysical questions about protein energy landscapes in native biological environments. One area of particular interest, explored in Chapter 4, is how protein folding is modulated by the ribosome during translation. As a field, we have learned an immense amount about the physical effect of the ribosome, and we now have an abundance of evidence that co-translational folding occurs regularly and can influence proper protein folding outcomes (outlined below). However, the majority of our experimental descriptions of nascent chain energy landscapes come from the study of nascent chains that are stalled on the ribosome. This eliminates a key component of co-translational folding, the vectorial nature of translation.

5.1.1 Folding during vectorial synthesis—a kinetic competition

For larger proteins, vectorial synthesis allows regions at the N-terminus the opportunity to fold before C-terminal regions have been synthesized.[1] This results in a complex kinetic competition between protein folding and translation where the energy landscape that a nascent chain can explore changes as each successive amino acid is added by the ribosome. Protein folding, and conformational changes in general, have a very broad distribution of timescales (see Chapter 1). Minor conformational changes, like transitions between side-chain rotamers, occur on the picosecond to nanosecond timescale, and more complex transitions, like the folding of entire domains, can be as fast as tens of microseconds and as slow as minutes.[2,3] Importantly, proteins that stably populate their native state, by definition, will unfold much slower than they fold. The range of protein unfolding rates for known two-state proteins ranges from hundreds of milliseconds to tens of days.[3] Protein translation rates fall in the middle of these timescales, where the addition of a single amino acid takes between 50 milliseconds and 1 second.[4] This means that proteins can explore regions of their conformational energy landscape, but not so thoroughly that equilibrium is necessarily reached, in other words, a kinetic competition occurs where certain unique structures can form during translation, and these structures can persist until after translation of the entire polypeptide is complete. This concept was elegantly explained by Clark as an evolving series of energy landscapes.[5] This raises the question - how can we

experimentally determine what the nascent chain energy landscape looks like for each translational intermediate, and how does the kinetic competition between translation and the exploration of these landscapes dictate protein-folding outcomes?

5.1.2 Current state of the nascent chain biophysics field

To date, the majority of detailed biophysical and structural studies of co-translational folding involve some mechanism of translational stalling. One of the most common approaches uses a specific amino acid sequence, known as a translational arrest peptide (the most commonly used being the SecM stalling sequence), that interacts strongly with the ribosome exit tunnel, halting translation and resulting in a stable ribosome-nascent chain (RNC) complex.[6,7] This arrest peptide approach allows for the purification of large quantities of homogeneous RNCs, facilitating the use of advanced biophysical methods.[8] This has allowed for the determination of nascent-chain structures by CryoEM[9–11], the study of nascent-chain dynamics and ribosomal interactions by NMR and HDX-MS[12–15], the quantitative measurement of global nascent-chain stability and kinetics with pulse proteolysis[16,17], and investigations of folding kinetics and pathways using single-molecule force spectroscopy[18,19], amongst many other experiments. These studies have been instrumental in understanding how protein energy landscapes are perturbed by the ribosome.[20,21] However, it is labor-intensive to perform these experiments for every possible nascent chain length, and in none of these experiments is the kinetic competition between translation and folding taking place.

Arrest peptides have also been leveraged, on their own, as effective probes for co-translational protein folding. When a domain in a nascent chain folds from a flexible disordered unfolded state to a more rigid globular state near the surface of the ribosome, it can generate a force that propagates through the ribosome exit tunnel.[22] This force dislodges the arrest peptide, allowing translation to continue.[6] The Von Heijne group, among others, has leveraged this behavior to develop force profile analysis (FPA).[9–11,22–25] In FPA many protein constructs are created, each with a stalling sequence, a protein of interest, and a variable length linking sequence in between the two. Each construct is then separately assayed to determine which lengths create enough force to release the translational stall. When this release occurs efficiently it is interpreted as a co-translational folding event. The relative ease of this approach allows the interrogation of many different proteins, revealing trends between size, charge, thermodynamic stability, and the propensity to fold near the ribosome.[22] However, even this experiment does not probe the true kinetic competition between translation and folding, as in each case the arrest peptide engages, and several minutes are allowed for any folding force to dislodge the arrest peptide.

Other methods more closely capture the true kinetic competition, but even they indirectly rely on stalling translation. For instance, ribosome profiling can map the points in translation when FKBP12-rapamycin binding domain (FRB) acquires a native-like structure, defined as its ability to bind the ligand rapamycin.[26] In theory, this approach probes folding events occurring during translation, without stalling sequences, and in the native cellular environment. However, the assay itself involves stalling since it requires the addition of small molecules to bind the ribosome and halt translation, allowing RNCs to be isolated for RNA footprinting. It is entirely possible that structures form during this extended time period that otherwise would not have had

time to form before translation continued. Another promising set of methods for monitoring structure formation during translation is FRET and PET, where one can monitor the distance between two probes in the nascent chain during the translation process. This is, of course, a technically challenging experiment since one has to incorporate two probes (either two fluorophores or a fluorophore and a quencher) into the actively translating nascent chain. Despite the challenges, co-translational FRET and PET have both provided an unprecedented level of detail about nascent chain structure formation during translation.[27–29] While this method does involve monitoring a probe during active translation, the deconvolution of the data to assign structure at the resolution of individual nascent chain lengths incorporates the FRET and PET values of stalled nascent chains at different lengths. Therefore, this method is also incapable of completely reproducing the kinetic competition between folding and translation.

Finally, there are several studies monitoring nascent chain folding during active translation without any stalling involved. One is described in chapter 4, where we used the ligand binding activity of the protein HaloTag to watch the acquisition of native structure during translation in real-time.[30] This experiment, along with other in vitro re-folding experiments, provided evidence that co-translational folding can prevent the formation of an aggregation-prone intermediate. This approach has several weaknesses, one being that it is not generalizable, as it depends on the activity and slow folding of HaloTag. Another weakness is that this assay reports on the folding of the protein after release from the ribosome, only indirectly reporting on events taking place during translation. Another conceptually similar approach was used by the Clark lab, where they cleverly designed a protein fusion construct based on two different fluorescent proteins, where a single beta-strand is shared between them.[31] As a result, each individual protein molecule can only successfully fold one of the two fluorescent protein domains. When this construct is translated, the N-terminal fluorescent protein is preferentially formed, providing direct evidence that the vectorial nature of translation can bias co-translational folding. Both of these studies provided convincing evidence that kinetic competition occurring during translation can bias folding outcomes, but neither provided detailed structural information on nascent chain folding during translation.

5.2 Future strategies for studying nascent chain folding during translation

What is needed are methods that can provide structural information of nascent chains, as ribosomes are actively synthesizing, without any stalling. The ideal method would (1) report on the structure of the nascent chain, either in terms of the acquisition of native structure or ideally more local information on the formation of secondary structures or burial of side chains (2) return this information as a function of nascent chain length and (3) do both of the above under conditions where the ribosome is actively translating the nascent chain. Here I propose several ways we can accomplish this.

5.2.1 Development of single-molecule approaches for monitoring co-translational folding in real-time

There are many detailed studies on the kinetics and mechanisms of translation using single-molecule fluorescence and single-molecule FRET.[32–34] By fluorescently labeling ribosomal subunits, tRNAs, initiation factors, elongation factors, and termination factors and monitoring their assembly and dynamics, we now have an extremely detailed view of nearly every assembly step and conformational change that takes place. These robust and powerful platforms can now be leveraged to ask questions about co-translational folding. Having already established that translation can be monitored in real-time for single ribosomes, all that is required is a fluorescence read-out that is sensitive to the folding of the nascent chain. In Chapter 4, we leverage the ability of HaloTag to rapidly, specifically, and irreversibly bind a fluorescent ligand. This could be adapted to a single-molecule approach, where a C-terminal linker could be appended to a protein, and during translation, as the domain exits the ribosome and folds, it would rapidly bind its fluorescent ligand indicating a folding event occurred. This could be combined with ribosomal subunit FRET and/or fluorescently labeled tRNAs to simultaneously monitor the addition of each amino acid to the nascent chain, and determine at exactly which nascent chain length folding occurred. The key problem with extending this experiment using HaloTag is that it folds extremely slowly. While this was an asset in Chapter 4, it means that in this experiment we would likely never observe HaloTag folding during translation. However, this same concept could be applied to any protein that rapidly binds a fluorescent molecule or a molecule that can be fluorescently labeled as long as the binding kinetics are suitably fast, and the dissociation is suitably slow.

Together with the Puglisi lab at Stanford University, we have already begun to make progress toward this goal. First, we chose other model systems where folding can be monitored by the binding of a molecule to the native state, *E. coli* Dihydrofolate Reductase (DHFR) which binds to Methotrexate (MTX), and an RNase from *Bacillus amyloliquefaciens* (Barnase) which binds to its inhibitor, the protein Barstar.[35,36] We demonstrated that by fluorescently labeling MTX and Barstar and monitoring the fluorescence polarization of these molecules during the translation of DHFR and Barnase, we can monitor the simultaneous translation and folding of these proteins (Fig. 5-1). This is the same experiment as the HaloTag fluorescence polarization experiment described in Chapter 4. However, since both DHFR and Barnase are known to fold rapidly, both likely fold immediately upon release from the ribosome. We verified this by halting the translation of DHFR with an antibiotic that binds to the ribosome. This quenching of the reaction immediately halted the increase in fluorescence polarization of MTX, confirming that folding occurred nearly immediately after the translation and release of the full-length nascent chain (Fig 5-1).

We went on to generate Barnase and DHFR constructs with SecM stall sequences and flexible linkers. We then demonstrated that we could capture individual ribosomes and translate DHFR and Barnase in the presence of fluorescently labeled MTX and Barstar. MTX proved to be too hydrophobic of a molecule for this single-molecule platform, sticking to the slide surface and providing a large background signal. The combination of Barnase and Barstar, however, proved that we could successfully monitor single ribosomes that could translate Barnase and that Barstar

could then bind to the folded stalled nascent chain (Fig. 5-1). The next steps for these experiments will be to couple this assay to existing assays for monitoring translation and tRNA binding, allowing us to simultaneously monitor both the folding of nascent chains and translation in real-time, thus providing the first codon level resolution of co-translational protein folding monitored in real-time.

This platform could be extended to ask various interesting biological questions. As monitoring the binding of proteins to the ribosome has been done for many elongation factors, it should be possible to similarly label and monitor the binding of co-translational chaperones like Trigger Factor and DnaJ/K[37–40], and co-translational modifying enzymes like peptide deformylase, methionine aminopeptidase, and N-terminal acetyltransferase.[41] Additionally, as nanobody design and de novo protein design continue to improve, it will likely become possible to develop bespoke probes of folding for any nascent chain.[42–44] Admittedly, the constructs that we have used so far are highly non-physiological, with artificial linkers and stalling sequences introduced. However, by generating binders that recognize non-linear structural motifs unique to the native (or non-native) state of a given protein, this same approach could be extended to any protein of interest. These binders would be fluorescently labeled and this same single molecule fluorescence approach would then determine at which nascent chain lengths a specific structural motif forms.

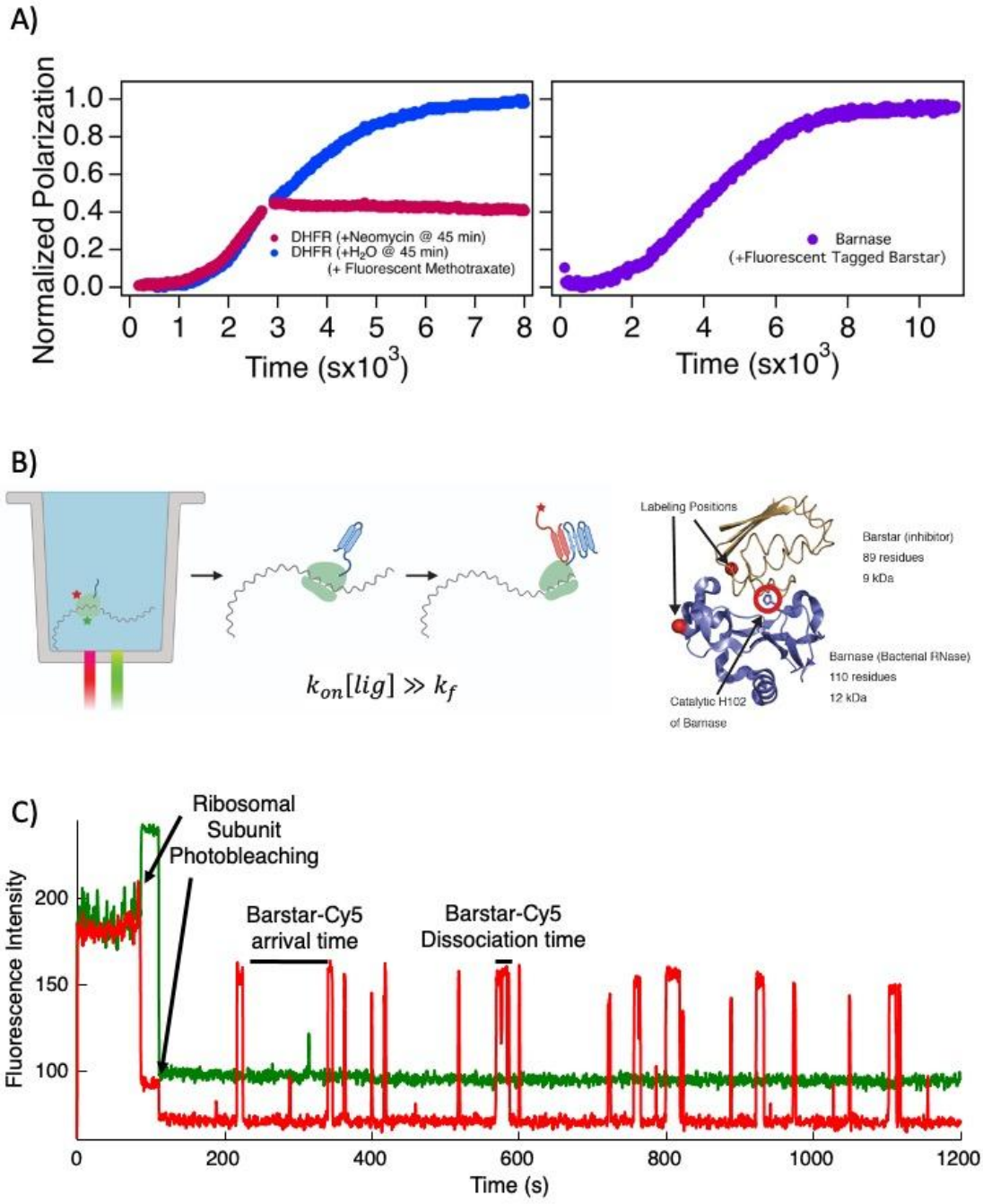


Figure 5-1. Monitoring co-translational folding with single-molecule fluorescence. (A) in vitro translation of the proteins DHFR (left) and Barnase (right), monitored by the polarization of fluorophore-modified ligands MTX(left, blue) and Barstar (right, purple). If translation is halted by the addition of neomycin (left, red), the increase in polarization is instantly stopped, indicating folding is occurring immediately after release from the ribosome. (B) Shown on the left is a diagram of a zero-mode waveguide (ZMW) well with an immobilized ribosome (green) attached to RNA (gray) with two fluorescent labels attached (green and red). As translation of the nascent chain (blue) continues, the nascent chain folds, and a fluorescent ligand binds (red). Shown on the right is the structure of the Barnase:Barstar complex, with the catalytic residue and labeling positions, indicated. (C) Sample FRET trace showing first photobleaching of the ribosome fluorophore labels, then the arrival of fluorescently labeled Barstar, indicating the complete translation and folding of the Barnase nascent chain.

5.2.2 Potential for covalent labeling methods to monitor co-translational folding

One weakness of the above approach is the probe is specific to attaining a native-like structure. A promising alternative approach would be to leverage the power of covalent labeling methods, such as those outlined in Chapters 1-3, as these provide local structural information on secondary structure or side chain burial. Additionally, as these methods are ideal for handling complex heterogeneous mixtures, they would be well suited to probe nascent chain structure formation during translation, either in *in vitro* settings or in the native biological environment. One experimental issue that plagues those who study co-translational folding is the synchronization of ribosomes.[45] As translation is a series of reactions, stochasticity will result in the desynchronization of the translation reaction, where various ribosomes will be at different stages of synthesis with variable-length nascent chains. One way to combat this is to monitor single ribosomes, as described in section 5.2.1. Another potential solution is to encode structural information into each nascent chain using a covalent labeling method like HDX or HRF, then separate nascent chains by length, and then read out the structural information with a method like mass spectrometry.

I have primarily focused on two covalent labeling methods in this dissertation, HDX-MS, and HRF-MS. HDX-MS has already been successfully used to study RNCs.[15] However, if our goal is to study nascent chains of variable length, it is unlikely that HDX-MS will be viable, for multiple reasons. First, most HDX-MS experiments rely on digesting proteins into peptides. If there is a heterogeneous mixture of nascent chain lengths, as we would expect based on the difficulty in ribosome synchronization, it would be impossible to know what nascent chain length a given peptide originated from. One way to overcome this would be to first separate nascent chains by length and then digest them into peptides.[46] This is unlikely to work for HDX-MS, as back exchange limits both the solution conditions (cold, pH~2.5) and the time between labeling and measurement (maximum of tens of minutes).[47] Given those restraints, it is unlikely that nascent chains could be first labeled then separated and digested without complete back exchange occurring. One alternative to both of the above approaches would be to bypass digestion entirely and only monitor intact nascent chains. This is an even more attractive approach when you consider that, as each nascent chain has a different mass, each would be able to be monitored separately by mass spectrometry. However, here too the restrictions imposed by back exchange likely prevent this strategy from working. Since each nascent chain is attached to a tRNA, measuring HDX by MS will likely be a challenge, due to the higher mass, the difficulty of ionization of such a negative molecule, and the ability of the tRNA itself to become deuterated. Similarly, it is unlikely that first removing the tRNA will be feasible, as the cold and acidic conditions will likely prevent timely enzymatic removal.

While HDX-MS will likely not be amenable to studying co-translational folding, HRF-MS does not suffer from the same back exchange problems described above. The oxidative modifications resulting from HRF are non-labile.[48] Meaning that the information encoded during the HRF labeling will remain stable during the various separation and digestion processes that will be required to study co-translational folding. Both of the approaches outlined above are likely feasible. After HRF labeling, all remaining RNA could be enzymatically degraded, and the degree of modification for each nascent chain could be directly measured by mass spectrometry

(Fig. 5-2). Once normalized using the known intrinsic rates (see Chapter 1) the resulting data could be plotted as a measure of protection (y-axis) as a function of nascent chain length (x-axis). If a co-translational folding intermediate buries side chains, protecting them from solvent, we would expect to see an increase in protection at the nascent chain length where that intermediate first forms. Additionally, higher structural resolution could be achieved by then separating nascent chains chromatographically, and then digesting and measuring the protection of individual peptides. In summary, the stability of the oxidation modifications in HRF makes HRF-MS a promising approach for studying nascent chain structure formation during translation.

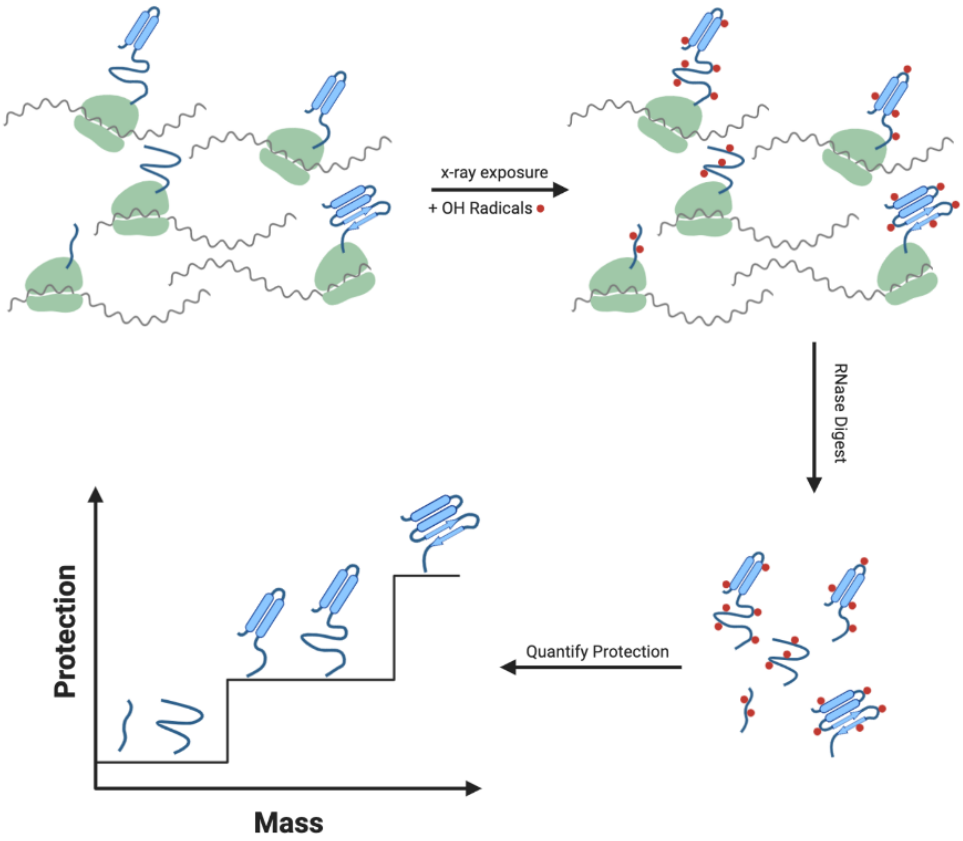


Figure 5-2. Detection of co-translational folding intermediates with HRF-MS. A heterogeneous mixtures of Ribosomes (green) translating RNA (gray) and extending the nascent chain (blue). At some timepoint the nascent chains can be labeled with hydroxyl radicals, and then the RNA and other proteins can be removed by digestion and purification. The oxidation of the mixture of labeled nascent chains can be measured with mass spectrometry. Once normalized to intrinsic rates of labeling, the protection at each nascent chain length can be determined, and increased protection should be observed for lengths that populate folding intermediates.

5.3 Conclusions

Every protein is translated by the ribosome, and the timescales of conformational changes suggest that proteins are extensively searching their energy landscapes during this process. Given the ubiquitousness of this process and the potential to dictate proper folding outcomes, the sequence determinants that govern this process must be further investigated. The complex kinetic interplay and the many macromolecular components involved have made descriptions of structures during translation very difficult to acquire. Here I suggest two possible methods that can describe either native state structure, or the formation of tertiary structure, by nascent chains at the resolution of individual translation elongation steps. These unprecedented structural descriptions of nascent chains will be instrumental in improving our understanding of co-translational folding.

5.4 Methods

5.4.1 Fluorescence polarization

Data collection: All experiments were performed at 37°C. Fluorescence polarization was performed on a BioTek Synergy Neo2 plate reader in 384-well black flat-bottom plates(Corning). Acquisitions were collected using polarizers as well as 530nm/590nm filters with side gain set at 45 and top gain set at 40. Read height was 7.5mm and 10 measurements were made per data point. After a 15-minute incubation at 37°C, readings were initialized by the addition of DNA (encoding either *E. coli* DHFR or *B. amyloliquefaciens* Barnase H102A D44C). Measurements were taken every 20 seconds for five hours after thirty seconds of mixing and a 90-second delay for temperature equilibration.

IVT Reactions: IVT reactions using the PURExpress system were set up on ice per the manufacturer's protocols for a 30µL reaction with the addition of 1µL RNase Inhibitor, Murine, and a final concentration of 10 uM fluorescently labeled ligand, either Fluorescien-Methotrexate (Thermo Fisher) or Fluorescein-Maleimide labeled Barstar C40A and pipetted into wells. Plates were covered with clear titer-tops to prevent evaporation and equilibrated at 37°C for at least 15 minutes. Reactions were initiated with 2µL of 125ng/µL plasmid DNA.

5.4.2 Single-molecule fluorescence

Single-molecule fluorescence assays were performed as previously described.[49,50] Briefly, single mRNA encoding for *B. amyloliquefaciens* Barnase H102A D44C, a $(GS)_{25}$ linker sequence, and a SecM stalling sequence were immobilized in zero-mode waveguide wells within a SMRT cell. Fluorescently labeled ribosome subunits were assembled and translation was initiated in the presence of translation components and Cy5-Maleimide labeled Barstar C40A. Fluorescence was monitored on a modified PacBio sequencer.

5.5 References

1. Zhang, G. & Ignatova, Z. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr. Opin. Struct. Biol.* **21**, 25–31 (2011).

2. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).

3. Maxwell, K. L. *et al.* Protein folding: defining a standard set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**, 602–616 (2005).

4. Young, R. & Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem. J* **160**, 185–194 (1976).

5. Clark, P. L. Protein folding in the cell: reshaping the folding funnel. *Trends Biochem. Sci.* **29**, 527–534 (2004).

6. Goldman, D. H. *et al.* Ribosome. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science* **348**, 457–460 (2015).

7. Bhushan, S. *et al.* SecM-Stalled Ribosomes Adopt an Altered Geometry at the Peptidyl Transferase Center. *PLoS Biol.* **9**, e1000581–10 (2011).

8. Cassaignau, A. M. E. *et al.* A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nat. Protoc.* **11**, 1492–1507 (2016).

9. Nilsson, O. B. *et al.* Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* **24**, 221–225 (2017).

10. Nilsson, O. B. *et al.* Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep.* **12**, 1533–1540 (2015).

11. Marino, J., von Heijne, G. & Beckmann, R. Small protein domains fold inside the ribosome exit tunnel. *FEBS Lett.* **590**, 655–660 (2016).

12. Chan, S. H. S. *et al.* The ribosome stabilizes partially folded intermediates of a nascent multi-domain protein. *Nat. Chem.* 1–9 (2022).

13. Cabrita, L. D., Hsu, S.-T. D., Launay, H., Dobson, C. M. & Christodoulou, J. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22239–22244 (2009).

14. Waudby, C. A. *et al.* Systematic mapping of free energy landscapes of a growing filamin domain during biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **11**, 201716252–201719749 (2018).

15. Wales, T. E. *et al.* Resolving chaperone-assisted protein folding on the ribosome at the peptide level. *bioRxiv* 2022.09.23.509153 (2022) doi:10.1101/2022.09.23.509153.

16. Samelson, A. J., Jensen, M. K., Soto, R. A., Cate, J. H. D. & Marqusee, S. Quantitative determination of ribosome nascent chain stability. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13402–13407 (2016).

17. Jensen, M. K., Samelson, A. J., Steward, A., Clarke, J. & Marqusee, S. The folding and unfolding behavior of ribonuclease H on the ribosome. *J. Biol. Chem.* **295**, 11410–11417 (2020).

18. Guinn, E. J., Tian, P., Shin, M., Best, R. B. & Marqusee, S. A small single-domain protein folds through the same pathway on and off the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12206–12211 (2018).

19. Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).

20. Waudby, C. A., Burridge, C., Cabrita, L. D. & Christodoulou, J. Thermodynamics of co-translational folding and ribosome-nascent chain interactions. *Curr. Opin. Struct. Biol.* **74**, 102357 (2022).

21. Kaiser, C. M. & Liu, K. Folding up and Moving on – Nascent Protein Folding on the Ribosome. *J. Mol. Biol.* (2018) doi:10.1016/j.jmb.2018.06.050.

22. Farías-Rico, J. A., Ruud Selin, F., Myronidi, I., Frühauf, M. & von Heijne, G. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9280–E9287 (2018).

23. Farías-Rico, J. A., Goetz, S. K., Marino, J. & von Heijne, G. Mutational analysis of protein folding inside the ribosome exit tunnel. *FEBS Lett.* **591**, 155–163 (2017).

24. Kemp, G., Nilsson, O. B., Tian, P., Best, R. B. & von Heijne, G. Cotranslational folding cooperativity of contiguous domains of α-spectrin. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 14119–14126 (2020).

25. Kemp, G., Kudva, R., de la Rosa, A. & von Heijne, G. Force-Profile Analysis of the Cotranslational Folding of HemK and Filamin Domains: Comparison of Biochemical and Biophysical Folding Assays. *J. Mol. Biol.* **431**, 1308–1314 (2019).

26. Han, Y. *et al.* Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12467–12472 (2012).

27. Mercier, E. & Rodnina, M. V. Co-translational Folding Trajectory of the HemK Helical Domain. *Biochemistry* **57**, acs.biochem.8b00293–3464 (2018).

28. Thommen, M., Holtkamp, W. & Rodnina, M. V. Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* **42**, 83–89 (2017).

29. Holtkamp, W. *et al.* Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104–1107 (2015).

30. Samelson, A. J. *et al.* Kinetic and structural comparison of a protein's cotranslational folding and refolding pathways. *Sci Adv* **4**, eaas9098 (2018).

31. Sander, I. M., Chaney, J. L. & Clark, P. L. Expanding Anfinsen's Principle: Contributions of Synonymous Codon Selection to Rational Protein Design. *J. Am. Chem. Soc.* **136**, 858–861 (2014).

32. Petrov, A., Chen, J., O'Leary, S., Tsai, A. & Puglisi, J. D. Single-molecule analysis of translational dynamics. *Cold Spring Harb. Perspect. Biol.* **4**, a011551–a011551 (2012).

33. Tsai, A., Puglisi, J. D. & Uemura, S. Probing the Translation Dynamics of Ribosomes Using Zero-Mode Waveguides. in vol. 139 1–43 (Elsevier, 2016).

34. Petrov, A. *et al.* Dynamics of the translational machinery. *Curr. Opin. Struct. Biol.* **21**, 137–145 (2011).

35. Rajagopalan, P. T. R. *et al.* Interaction of dihydrofolate reductase with methotrexate: Ensemble and single-molecule kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13481–13486 (2002).

36. Schreiber, G. & Fersht, A. R. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486 (1995).

37. Deckert, A. *et al.* Common sequence motifs of nascent chains engage the ribosome surface and trigger factor. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

38. Ferbitz, L. *et al.* Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).

39. Oh, E. *et al.* Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295–1308 (2011).

40. Wruck, F. *et al.* Protein folding mediated by trigger factor and Hsp70: New insights from single-molecule approaches. *J. Mol. Biol.* (2017) doi:10.1016/j.jmb.2017.09.004.

41. Yang, C.-I., Hsieh, H.-H. & Shan, S.-O. Timing and specificity of cotranslational nascent protein modification in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23050–23060 (2019).

42. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017).

43. Chevalier, A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).

44. Schoof, M. *et al.* An ultra-high affinity synthetic nanobody blocks SARS-CoV-2 infection by locking Spike into an inactive conformation. *Cold Spring Harbor Laboratory* 2020.08.08.238469 (2020) doi:10.1101/2020.08.08.238469.

45. Rodnina, M. V. & Wintermeyer, W. Rapid Kinetic Analysis of Protein Synthesis. in 119–139 (Springer New York, 2012).

46. Mayne, L. *et al.* Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation-mass spectrometry method. *J. Am. Soc. Mass Spectrom.* **22**, 1898–1905 (2011).

47. Engen, J. R. & Wales, T. E. Analytical Aspects of Hydrogen Exchange Mass Spectrometry. *Annu. Rev. Anal. Chem.* **8**, 127–148 (2015).

48. Xu, G. & Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**, 3514–3543 (2007).

49. Prabhakar, A. *et al.* Uncovering translation roadblocks during the development of a synthetic tRNA. *Nucleic Acids Res.* **50**, 10201–10211 (2022).

50. Prabhakar, A., Capece, M. C., Petrov, A., Choi, J. & Puglisi, J. D. Post-termination Ribosome Intermediate Acts as the Gateway to Ribosome Recycling. *Cell Rep.* **20**, 161–172 (2017).