Title

Probing Protein Energy Landscapes Under Physiologically Relevant Conditions

Author

Volz, Sara

Publication Date

2023

Peer reviewed|Thesis/dissertation

Probing Protein Energy Landscapes Under Physiologically Relevant Conditions

By

Sara E Volz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Susan Marqusee, Chair
Professor Andreas Martin
Professor David Savage
Professor Ahmet Yildiz

Fall 2023

Abstract

Probing Protein Energy Landscapes Under Physiologically Relevant Conditions

by

Sara E Volz

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Susan Marqusee, Chair

It is imperative to investigate protein energy landscapes within the framework of physiological or near-physiological conditions, particularly when emphasizing the significance of the protein folding/unfolding process or of specific unfolded states in relation to a protein's biological function.

Ubiquitination, a post-translational modification instrumental in proteasomal degradation of protein substrates, can influence protein stability. Proteins experience force-mediated denaturation when unfolded by the proteasome; therefore, I employed single-molecule force spectroscopy to study rates of unfolding of ubiquitinated proteins. I demonstrate that ubiquitination of the model protein barstar at lysine 60 not only increases its force-induced unfolding rate, but also influences the responsiveness of unfolding rate to force, implicating a change in unfolding pathways.

Human γD-crystallin is a monomeric protein abundant in the eye lens nucleus that must remain stably folded for an individual's entire lifetime to avoid aggregation and cataract formation. It is not clear whether intermediates of γD-crystallin populated under denaturing conditions are also populated under native conditions, or what relevance they hold to the mechanism of aggregation. Therefore, I employed hydrogen-deuterium exchange coupled with mass spectrometry (HDX-MS) to characterize several variants of γD-crystallin under both native and denaturing conditions. By investigating several types of mutations in each of γD-crystallin's two domains, I show that, for two cataract-prone variants of γD-crystallin, the lowest-energy equilibrium intermediate populated under native conditions is structurally and energetically distinct to the intermediate populated under chemically denaturing conditions. The interface between the two domains is crucial to the formation of this new intermediate, and disruption of the interface either by mutation or by mild denaturation permits direct observation of both intermediates at the same time. The natively populated partially folded conformation of γD-crystallin exposes a surface which is normally buried both in the full-length structure and in isolated folded domains. Together with the noted effects of mutations, this suggests it may be significant to aggregation and to cataract formation.

Table of Contents

Chapter 1. Introduction 1

1.1 Energy landscapes are the determinants of biological protein function 1

1.2 High-energy states on the energy landscape are instrumental in protein folding, function, and misfunction 2

1.3 Partially folded states are complicated to study 3

1.4 Scope and Summary 4

Chapter 2. Characterizing Intermediate States of Human γD-Crystallin under Native Conditions 5

2.1 Introduction 5

2.1.1 Crystallin aggregation in the eye lens leads to cataract formation 5
2.1.2 The energy landscape of γD-crystallin is implicated in aggregation 6
2.1.3 Hydrogen-deuterium exchange mass spectrometry can detect and characterize rare states in the native ensemble 7
2.1.4 Scope and summary 8

2.2 Results 9

2.2.1 Equilibrium unfolding/refolding of wild-type and mutant γDC 9
2.2.2 Continuous-labeling HDX-MS on variants of γD Crystallin 14
2.2.3 γDC hydrogen-deuterium exchange is slow at the interface 16
2.2.4 HDX in isolated domains of γDC 19
2.2.5 Disruption of the interface in the CTD destabilizes the native state intermediate 21
2.2.6 Peptides with EX1 kinetics report on a correlated opening reaction suggestive of slow interconversion between two states 24

2.3 Discussion 27

2.3.1 Ising analysis quantifies interdomain interaction energetics 27
2.3.2 The native-state intermediate is distinct from the denatured intermediate 28
2.3.3 Mutation and denaturant skew relative population of the two intermediates 29

2.4 Conclusions 30

2.5 Materials and Methods 32

2.5.1 Expression and purification of γD-crystallin variants 32
2.5.2 Determination of global stability by intrinsic tryptophan fluorescence 33
2.5.3 HDX-MS continuous exchange 36
2.5.4 HDX-MS data analysis 37

Chapter 3. Probing the mechanical unfolding of a ubiquitinated protein using optical tweezers 41

3.1 Introduction 41

3.1.1 Ubiquitin is a post-translational modification necessary for proteasomal degradation 41
3.1.2 Force is a biologically relevant perturbant 43

3.1.3 Single-molecule force spectroscopy can be used to exert and measure piconewton-scale forces on molecules 44
3.1.4 Ubiquitin attachment proceeds via an enzymatic cascade 45
3.1.5 Scope and summary 46
3.2 Handle attachment strategies 47
3.2.1 Negative result: sortase 49
3.2.2 Positive result: GMBS with short handles 50
3.3 Results 51
3.3.1 Force-extension curve analysis 52
3.3.2 Force-jump analysis 54
3.4 Conclusions, limitations, and future directions 59
3.5 Methods 60
3.5.1 Purification of barstar substrates 60
3.5.2 Reductive methylation of ubiquitin 61
3.5.3 Purification of ubiquitination enzymes 62
3.5.4 Preparation of homogenous mono-ubiquitinated substrate proteins 62
3.5.5 Preparation and attachment of DNA handles to protein substrates 63
3.5.6 Optical trap experimental setup 65
3.5.7 Data analysis for kinetics and for force jumps 65
References 66

Acknowledgements

First, I'd like to thank all my lab mates and colleagues over the years for their support, advice, and camaraderie, scientific or otherwise. Collecting data in the middle of a global pandemic was hard enough and would have been impossible without the support and collaboration of this wonderful lab. Thanks especially to Dr. Emma Carroll for her invaluable mentorship and instrumental guidance on the ubiquitination project, and a special shout-out to Bay Gamma—always ready with pet pictures or New York Times games for distractions when the science isn't cooperating.

I'd also like to thank my advisor, Prof. Susan Marqusee, for her invaluable mentorship and support. Thank you especially for always being ready to suggest new projects, ideas, and directions when I felt stuck. The culture of this lab—always asking questions and getting deep into the weeds of protein thermodynamics at lab meetings, always ready to support everyone's projects and ideas—is very special, and I'm very grateful for how you foster this spirit of collaboration and scientific creativity.

Thanks to my friends: the whole D&D crew who let me GM a five-year journey with them from level one to level twenty, and everyone who comes to karaoke nights or twelve-hour board game sessions at Victory Point.

Thank you to my partner, Amanda Jack: thank you for being by my side on this journey, from making fun of bad musicals to appreciating good ones, from hot-air ballooning to catnip growing, from drain-fly battling to Comic Con adventuring, from board-game collecting to (reluctant) hiking. Can't wait to see where we go next.

Thanks to my family: my mom, Pattye Volz, who's always been my biggest advocate in every goal I've ever had, from musical theater to science fair (even if she had to help me write the emails); my dad, David Volz, who encouraged my interest in science from a young age (even if it involved cat harassment); and my brother, Caleb Volz, for spurring me in the ways only an older brother can (i.e. needling). My family have always made me believe that I could accomplish anything I set out to do.

Finally, thanks to my not-really-a-kitten-anymore Caboodle, for being orange, soft, and bewhiskered.

Chapter 1. Introduction

1.1 Energy landscapes are the determinants of biological protein function

Proteins are polymers of amino acids that fold to produce biological macromolecular devices capable of an incredibly broad range of functions. How can a protein's linear amino acid sequence encode such a wide diversity of protein structure and function? Recent breakthroughs in machine learning have enabled remarkably accurate predictions of the three-dimensional structure corresponding to a given primary sequence (1). However, a protein's sequence defines more than just its native structure; it also defines an energy landscape encompassing many conformations with a complex array of energetics and dynamics. Consequently, it is imperative not to perceive proteins as static structures with one unchanging shape, but rather as statistical ensembles with the ability to populate many conformations. Achieving a comprehensive prediction of not just protein structure but also protein function from sequence necessitates a profound understanding of the intricate energy landscapes governing these biomolecules.

In order to function, proteins must navigate this conformational energy landscape both to fold to their native state and carry out their specific activities. Such landscapes are often depicted as “folding funnels”, which are multidimensional mappings illustrating possible conformations that a protein can adopt, organized by the free energy of each conformer, as well as the barriers between those conformations (2) (Figure 1.1). The classic studies of Anfinsen (3) and Levinthal (4) demonstrated that a small protein can fold spontaneously to its native state on a reasonable timescale, despite the impossibly large theoretical search space in the full conformational landscape. These findings indicate that a protein’s energy landscape is encoded in the protein’s amino acid sequence, and that protein folding is an ordered or biased reaction rather than a random search.

Within this energy landscape, the thermodynamic stability of a protein is characterized by the depth of its energy minima, while the kinetic stability is determined by the heights of the barriers separating these minima. The width of each minimum is linked to the diversity of conformations within the energy well. Many competing weak interactions define the protein's energy landscape and shape the search for its native fold. These interactions also facilitate conformational changes, allowing proteins to utilize malleability for applications such as binding-induced folding, enzymatic function, and allosteric regulation (5, 6). Consequently, the features of energy landscapes are shaped by the necessities for proteins to carry out their roles. Moreover, the evolution of protein function imposes constraints on the energy landscapes, often competing with the optimization of stability and function (7). Thus, the evolutionary drive to adopt landscapes with minimal frustration is balanced by the necessity for proteins to carry out their functions (8).

1.2 High-energy states on the energy landscape are instrumental in protein folding, function, and misfunction

Depending on the shape of the landscape, many partially folded states may be populated between the native ensemble at the bottom of the funnel and the unfolded ensemble at the top. Excursions from the native basin are interesting and pertinent to both function and malfunction, and both high-energy states involved in disease and functional transformations may be populated under some cellular conditions. Therefore, misfolded and partially or even fully unfolded protein states often have essential biological relevance.

For example, partially folded protein conformations can play a significant role during the initiation of the protein folding process in the cell. During protein biosynthesis, unfolded and partially folded conformers may be populated while the protein is still being translated by the ribosome, as the rate of folding is generally faster than the rate of translation (9).

A certain degree of structural flexibility is also considered essential to many types of protein-protein and protein-ligand interactions, including enzymatic catalysis. Active site accessibility, substrate binding and orientation, and stabilization of intermediates, among other functions, are all influenced by a protein's dynamic nature (10). Even more dramatically, some natural proteins can undergo extreme conformational changes under native conditions, sometimes even involving massive rearrangement of secondary structure (11).

Moreover, partial unfolding that exposes hydrophobic surfaces buried in the native state has been implicated in aggregation (12). Aggregation of misfolded or unfolded proteins has been identified as a core cause of more than twenty different so-termed "conformational diseases" in humans, including neurodegenerative disorders such as Parkinson's Disease (13). Aggregation can be considered as a major side-reaction of protein folding, with folding intermediates of particular relevance as they may be precursors to both native and misfolded ensembles (14). Mechanisms of aggregation occur through dynamic pathways of structural interconversion and demand in-depth characterization (Figure 1.1).

Finally, unfolding and/or partial unfolding is involved in the end of a protein's life cycle. Prior to proteolysis of proteins targeted for degradation, e.g. by the 26S proteasome in eukaryotes or by ClpXP in bacteria, proteins must be partially unfolded for engagement by the ATPase motors within these proteolytic machines (15). Hence, transient excursions to partially unfolded states are highly relevant to this process.

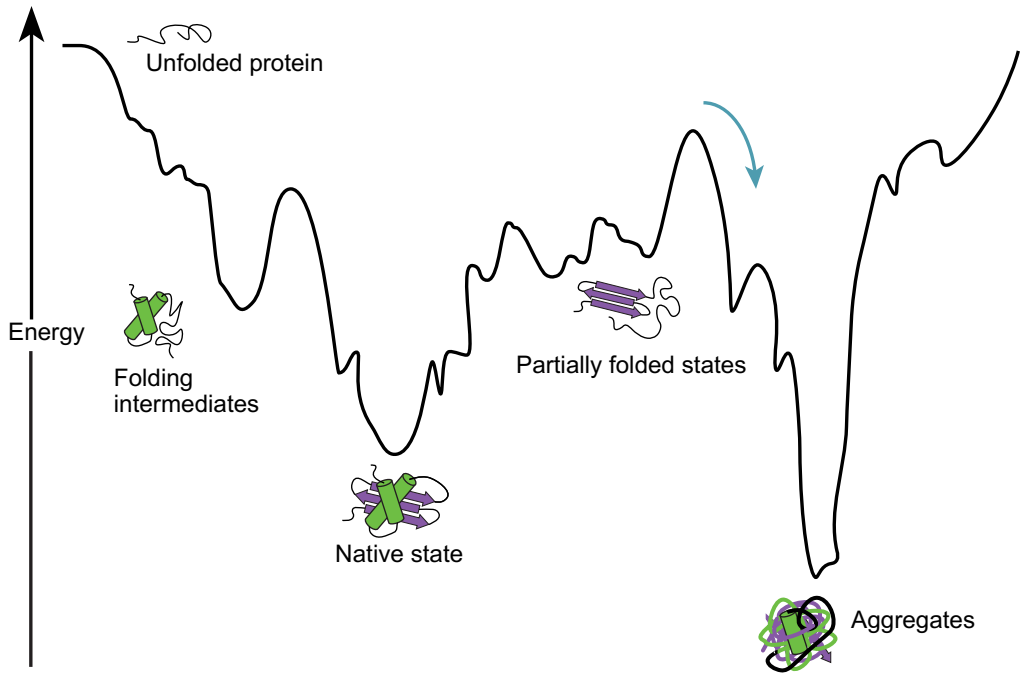


Figure 1.1: Schematic of a protein energy landscape. Polypeptides fold by sampling various conformations in the many-dimensional folding funnel. Local energy minima, which are partially folded intermediates, may be on- or off-pathway. Some partially folded states may have a propensity to form aberrant intermolecular or intramolecular contacts and thereby be at risk of misfolding or aggregation. Adapted from Hartl *et al.* (16).

1.3 Partially folded states are complicated to study

There are two major complicating factors to the study of these partially folded states. One is that a protein's sequence is not the only determinant of its energy landscape: the shape of the landscape is highly sensitive to changes in environment. Crucially, *in vitro* studies of the mechanisms of protein folding and unfolding may fail to recapitulate critical details of the dynamics or folding process *in vivo* (16, 17). Chemical-induced denaturation is not always the proper proxy for predicting the equilibrium behavior of a protein in its native environment in the cell, and the intrinsic properties of the polypeptide chain and the extrinsic factors of the cellular environment both dictate which intermediate and partially folded states are favored on the energy landscape.

Another complication is that intermediates are by their nature difficult to capture and study, especially by traditional structural or energetic methods. When involved in transformations – such as the protein folding process – intermediates are inherently ephemeral. Moreover, under native conditions, these states are rarely occupied, making it also difficult to detect high energy partially folded states at equilibrium. Sometimes, one can perturb the system from native conditions (e.g. by temperature or chemical

denaturation) to bias the energy landscape towards a partially folded state in order to observe it, followed by extrapolation back to native conditions. Such models assume that the ensemble of partially structured states observed under denaturing conditions is the same as that populated under native conditions, which may not always be the case.

Thus, there is a compelling need to study protein energy landscapes in the context of physiological or near-physiological conditions, especially when invoking the relevance of the protein folding/unfolding process or of particular unfolded states to a protein's biological function.

1.4 Scope and Summary

Herein, I present the work completed for my PhD towards the goal of uncovering hidden features in energy landscapes under physiologically relevant conditions.

First, I present my work aimed at identifying and characterizing intermediates associated with the energy landscape of γD-crystallin, a protein found in the eye lens which can cause cataract formation upon aggregation. By employing both hydrogen-deuterium exchange/mass spectrometry and traditional chemical denaturation, I sought to a) identify the structural and energetic characteristics of γD-crystallin's equilibrium intermediate under non-denaturing conditions, and b) show how altering interfacial interactions between γD-crystallin's two domains can change those characteristics.

I discovered that, for two cataract-prone variants of γD-crystallin, the primary equilibrium intermediate populated under native conditions is structurally and energetically distinct to the intermediate populated under chemically denaturing conditions. I also demonstrate that the interface between the two domains is crucial to this new intermediate and show that disruption of the interface either by mutation or by mild denaturation permits direct observation of both intermediates at the same time. This newly identified partially folded conformation of γD-crystallin may be significant to aggregation and to cataract formation. These studies serve as an illustration of the importance of studying energy landscapes under conditions that mimic physiological ones, as intermediate states characterized under denaturing conditions are not always relevant to native conditions.

Second, I describe my efforts to characterize force-induced unfolding trajectories of ubiquitinated proteins by the use of single-molecule force spectroscopy. Ubiquitinated proteins experience force-mediated denaturation when unfolded by the proteasome, so I sought to use a similar context to study their rates of unfolding. My goal was to establish a means to study ubiquitinated proteins using optical trapping, and then to investigate whether ubiquitin can affect how a protein's energy landscape is perturbed by force. I found that ubiquitination of a single-lysine substrate not only increases its rate of mechanical unfolding, but also potentially changes the pathway by which the substrate unfolds.

Chapter 2. Characterizing Intermediate States of Human γD-Crystallin under Native Conditions

2.1 Introduction

Human γD-crystallin is a monomeric protein abundant in the eye lens nucleus that must remain stably folded for an individual's entire lifetime to avoid aggregation and cataract formation. γD-crystallin's two homologous domains, the N-terminal domain (NTD) and the C-terminal domain (CTD), interact via a strong hydrophobic interface. Several familial NTD-destabilizing mutations are linked to early-onset cataract, including V75D. When unfolded via chemical-induced denaturation *in vitro*, V75D γD-crystallin populates an intermediate at moderate levels of denaturant with the NTD completely unfolded and the CTD native-like. However, chemical-induced denaturation is not always the proper proxy for predicting the behavior of a protein in its native environment in the cell, and it is not clear whether this intermediate is populated under non-denaturing conditions, or what relevance it has to aggregation mechanisms.

I employed hydrogen-deuterium exchange mass spectrometry (HDX-MS) to characterize several variants of γD-crystallin under native and denaturing conditions. I investigated destabilizing mutations in the NTD, interface-weakening mutations in the CTD, and individual domains of γD-crystallin in isolation. For two cataract-prone variants of γD-crystallin, V75D and W42R, the primary equilibrium intermediate observed under native conditions retains partial structure in the NTD and is thus structurally and energetically distinct from the intermediate populated under moderately denaturing conditions, in which the NTD is fully unfolded. Importantly, the NTD-CTD interface is crucial to the formation of this new intermediate, and disruption of the interface either by mutation or by mild denaturation destabilizes this conformation and permits direct observation of both intermediates at the same time.

This newly identified intermediate, which is the most-populated partially folded conformation under native conditions in two cataract-prone variants of γD-crystallin, exposes a surface which is normally buried both in the full-length protein and in the protein's isolated folded domains. This intermediate may be significant to aggregation and to cataract formation. These results illustrate the importance of studying energy landscapes under conditions that mimic physiological ones, as the energy landscape under denaturing conditions cannot always be faithfully extrapolated to native conditions.

2.1.1 Crystallin aggregation in the eye lens leads to cataract formation

The lens of the eye is a unique and highly specialized organ subject to specific biophysical constraints and characteristics. It must be transparent so as to prevent light scattering or absorption, it must possess a high refractive index to tightly focus gathered light onto the retina, and its focal length must be adjustable to permit focusing at different distances (18, 19).

To achieve these optical properties, fully differentiated lens fiber cells consist of ~90% soluble, densely packed crystallin proteins at concentrations exceeding 400 mg/ml (20). During lens maturation, lens cells eliminate cellular structures that could contribute to light scattering, including nuclei and organelles, resulting in essentially no protein turnover at the core of the mature lens (21, 22).

Effectively, this means that crystallins must remain soluble and stably folded for an individual's entire lifetime. When damaged or partially unfolded crystallins aggregate, insoluble precipitates can form cataracts, an opacification of the eye lens which, with more than 20 million people affected, is the leading cause of blindness worldwide (23).

Crystallins are subdivided into three families: α-, β-, and γ-crystallins. α-crystallins are small heat shock proteins that assemble into oligomeric structures that perform chaperone functions and assist in homeostasis of β- and γ-crystallins (24). β- and γ-crystallins, which exist as dimers/oligomers and monomers, respectively, are small, highly stable proteins with high solubility that play structural roles in conferring the optical properties of the eye lens (25).

2.1.2 The energy landscape of γD-crystallin is implicated in aggregation

Human γD-crystallin (γDC) is the second-most abundant γ-crystallin found in the lens nucleus, comprising ~11% by mass of the total lens protein in young human lenses (26). It is relatively small (~21 kDa) and is comprised of two domains connected by a short linker, the N-terminal domain (NTD) and the C-terminal domain (CTD) (27). Each homologous domain contains two antiparallel β-sheet arrangements known as Greek keys, with strong nonpolar covalent interactions constituting the interface between the two domains (Figure 2.1A).

Despite the high thermodynamic and kinetic stability of γDC, chemical insult, such as UV damage (28), deamidation (29), or oxidation (30), can result in aggregation leading to cataract formation and blindness (31). Furthermore, mutations in the NTD of γD-crystallin have been linked to early-onset or congenital cataract, including R14C (32), P23T (33, 34), W42R (35) , R58H (36), G60C (37), V75D (38), R76S (39), and I81M (40), among others.

Interestingly, all of the mutations just listed, including V75D (41) and W42R (42), do not cause large-scale conformational perturbation or misfolding of the native state, demonstrating that differences in folding/unfolding behavior (that is, the energy landscape) are likely to be the source of aggregation-prone behavior instead. It has been proposed that small populations that exhibit partial unfolding behavior and are difficult to detect in solution may serve as the nucleating or aggregating species that leads to cataract (12).

In order to better understand the mechanism of aggregation and cataract formation, γDC and many of its variants have been well-characterized by equilibrium and kinetic

unfolding/refolding experiments using chemical denaturants (43–46). Wild-type γDC populates an equilibrium intermediate when chemically denatured by guanidinium chloride (GdmCl); this intermediate is best modeled as a fully unfolded NTD and a fully folded CTD (44, 46–48).

The interface between the N- and C-terminal domains of γDC is known to be crucial to its folding and stability. Isolated N-terminal and C-terminal domain fragments of γDC (hereafter termed γDC_{NTD} and γDC_{CTD}) are stable and well-folded in isolation. However, γDC_{NTD} is less stable than γDC_{CTD} despite the two equilibrium unfolding transitions being nearly equivalent in free energy in the full-length protein (44, 49). Therefore, in the context of the full-length protein, interactions at the domain interface contribute interfacial stability to the otherwise less-stable NTD. Past work has estimated the contribution of the interface to NTD stability to be approximately 4 kcal/mol (49), and destabilization of this interface, as in the variant V132A in the CTD, results in destabilization of the NTD unfolding transition (45).

Destabilization of the NTD through mutation, as in the mutant V75D, results in population of γDC's equilibrium intermediate across a broad range of denaturant concentrations (41). This intermediate has been structurally characterized by both NMR and SAXS for the V75D variant and is comprised of a fully unfolded NTD and a fully folded CTD (50). The direct structural observation of this equilibrium intermediate raises the tempting possibility that this partially unfolded species might serve as a specific nucleus for aggregation and/or cataract formation. However, the strength of γDC's domain interface, which is comparable to the intrinsic stability of the entire NTD, questions whether or not this intermediate would be the most accessible species under native conditions. I therefore sought to investigate both a) structural and energetic characteristics of γDC's equilibrium intermediate under non-denaturing conditions and b) how altering interfacial interactions between the NTD and the CTD can change those characteristics.

2.1.3 Hydrogen-deuterium exchange mass spectrometry can detect and characterize rare states in the native ensemble

Under native conditions, that is, in the absence of denaturant, partially folded species are rare and difficult to probe using traditional structural methods. Traditional means of measuring the conformational stability of a protein under native conditions rely on thermal or chemical denaturation, followed by analyzing the data using a two-state ($N \rightleftharpoons U$) or three-state ($N \rightleftharpoons I \rightleftharpoons U$) linear-extrapolation model to determine the free energy in the absence of denaturant (51). In other words, this model relies on an extrapolation of behavior under denaturing conditions to behavior in native conditions. However, the ensemble of partially structured states populated under denaturing conditions may not be identical to the ensemble accessible under native conditions. Therefore, I turned to hydrogen–deuterium exchange coupled with mass spectrometry (HDX-MS) to characterize the conformational and energetic landscape of γDC and various mutants under native conditions. While traditional equilibrium melts cannot detect states with

small populations under non-denaturing conditions, HDX-MS can directly detect high-energy states in the native ensemble.

HDX-MS measures exchange of amide hydrogens located in the protein backbone with deuterons in the solvent (52). In order for an exchange event to be successful, an amide proton must be exposed to the deuterated solvent via some opening transition. Observed HDX rates for a given peptide are dictated by the accessibility of amide protons within that peptide; therefore, faster-exchanging peptides are associated with regions in the protein experiencing higher solvent exposure and flexibility, while slower-exchanging peptides may contain amide hydrogens that are buried within the protein or involved in hydrogen bonds.

The kinetic and thermodynamic relationships that connect structural unfolding events with the observed rate of H/D exchange have been well-characterized according to the Linderstrøm-Lang equation, closed ⇌ open → exchanged (53). The observed exchange rate may report on either the equilibrium constant or the rate of opening of the unfolding event observed, depending on the kinetic regime involved (so-called EX1 versus EX2 behavior). HDX-MS is therefore an ideal tool to characterize systems like γDC, a protein for which sub-global unfolding events under native conditions are particularly germane to function and misfunction.

2.1.4 Scope and summary

Here, I have used HDX-MS to characterize partially unfolded intermediate states of γD-crystallin under both native and denaturing conditions. I find that the intermediate detected by HDX-MS under partially denaturing conditions is comprised of, as previously described, a folded CTD and completely unfolded NTD. However, I also find that, for two cataract-prone variants of γDC with destabilizing mutations in the NTD, the most accessible partially-unfolded state under native conditions is not this previously identified intermediate, but rather a distinct species with a folded CTD and partial structure at the NTD/CTD interface. Moreover, disruption of the interface, either by mutation or by mild denaturant, destabilizes this newly identified conformation, permitting direct observation of both intermediates at once. This newly identified conformation, which is more populated in γDC with familial mutations known to be involved in cataract formation, exposes a surface that is normally buried both in the full-length structure and in isolated folded domains. This intermediate may therefore be important along the pathway to aggregation and cataract formation.

2.2 Results

2.2.1 Equilibrium unfolding/refolding of wild-type and mutant γDC

Although many variants of γDC have already been well-characterized by chemically induced denaturation, all were analyzed by standard two- or three-state equilibrium unfolding models with linear extrapolations (44, 45, 47, 54). Therefore, I repeated these studies and analyzed the data using an Ising model to calculate interaction energies with more precision (55). Using an Ising analysis to generate a model for the interfacial contribution to protein stability requires equilibrium unfolding curves for the isolated single-domain fragments and full-length counterparts for each variant.

To monitor global equilibrium unfolding, I used GdmCl-induced denaturation monitored by tryptophan fluorescence (Figure 2.1B). γDC contains four tryptophan residues, two in the N-terminal domain (residue numbers 42 and 68) and two in the C-terminal domain (residue numbers 130 and 165). Since all four tryptophans are buried in the native structure and hence their fluorescence is highly quenched when γDC occupies the native state, intrinsic tryptophan fluorescence serves as an excellent reporter of folding in γDC (56, 57).

V75D is a cataract-causing NTD mutation known to heavily destabilize the NTD (58). A series of systematic mutational studies on interdomain interactions of γDC by Flaugh et al. revealed the contributions of crucial interface residues to both stability and folding rates of γDC, identifying residues with particular impact on the interaction energy (45, 54). One such mutation, V132A, is involved in a critical hydrophobic cluster with M43, F56, and I81 in the NTD.

Variants examined were wild-type γDC, V132A γDC, V75D γDC, V75D V132A γDC, γDC with five interface-involved mutations in the CTD, V132A, Q143A, L145A, M147A, and V170A (dInt-γDC), the isolated NTD fragment (γDC_{NTD}, residues 1-81), the isolated CTD (γDC_{CTD}, residues 84-174), and the isolated CTD with V132A (V132A-γDC_{CTD}). GdmCl-induced equilibrium denaturation was monitored by Trp-fluorescence (Figure 2.1). Data were analyzed via either a two state ($N \rightleftharpoons U$) or three-state ($N \rightleftharpoons I \rightleftharpoons U$) linear extrapolation model, resulting in parameters describing ΔG in the absence of denaturant, the dependence of ΔG upon denaturant concentration (*m*-value), and the folded and unfolded baseline intercepts and slopes for each transition (59, 60).

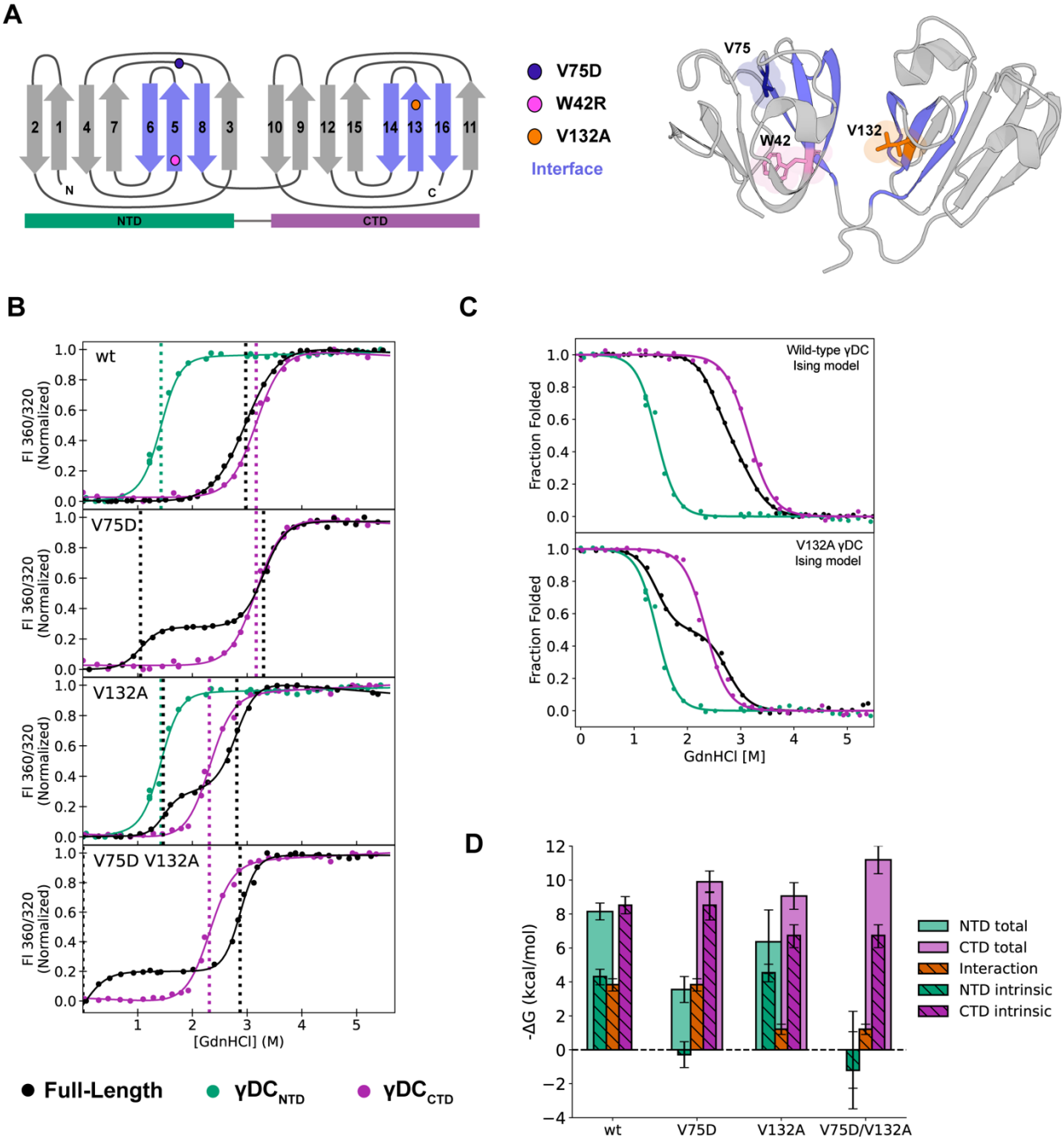


Figure 2.1: Equilibrium denaturation experiments of human γD-crystallin variants. A) Topological map and crystal structure (1HK0 (27)) of γDC, with Greek key beta strand architecture, NTD and CTD, and several mutations indicated. The interface between the two domains is highlighted in blue. B) Two- and three-state fits to tryptophan fluorescence data of full-length constructs and their corresponding isolated individual domains (γDC_{NTD}, green; γDC_{CTD}; purple; full-length γDC, black). C_M values are indicated with dotted lines to draw the eye. C) Global Ising fits from analysis of the same data for V132A and wild-type (γDC_{NTD}, green; γDC_{CTD}, purple; full-length γDC,

black). D) Graphical estimation of the relative energetic contributions of each domain's intrinsic stability and interfacial interaction energy to each transition (NTD, green; CTD, purple; interface, orange). Total transitions (solid bars)—that is, transitions calculated irrespective of dissection into component intrinsic and interaction energies—are taken from three-state fits, except in wild-type, where a three-state fit could not be obtained, and so the Ising interface and intrinsic γDC_{NTD} energies were summed for the total NTD transition. Intrinsic energies (striped bars) are based on values from Ising fits. In the case of constructs containing V75D where Ising fits could not be obtained, the corresponding interaction energy in wild-type or V132A was subtracted from the construct's three-state NTD transition to estimate the intrinsic energy of V75D γDC_{NTD} and V75D/V132A γDC_{NTD}, respectively, under the assumption that the V75D mutation does not impact the interaction energy. Parameters for all fits may be found in Table 2.1. Error bars represent standard error of the fit. Data were collected at 25˚C in PBS pH 7.0, 5 mM DTT.

Wild-type γDC and all isolated domains (γDC_{NTD}, γDC_{CTD}, and V132A-γDC_{CTD}) show cooperative apparent two-state behavior consistent with previously reported equilibrium data (44, 49) and were therefore analyzed using the two state ($N \rightleftharpoons U$) model (Figure 2.1B). Transition midpoints (C_m's) were calculated to be 3.0, 1.4, 3.2, and 2.3 M GdmCl, with corresponding *m*-values of 2.0, 3.4, 2.5, and 2.6 kcal mol^{-1} M^{-1} and free energies of -5.9, -4.8, -7.8, and -6.2 kcal/mol, respectively (Table 2.1).

All full-length crystallin constructs studied with the exception of wild-type γDC fit well to a three-state model ($N \rightleftharpoons I \rightleftharpoons U$), indicating the presence of an equilibrium intermediate. V75D, V132A, V75D/V132A, and dInt all show a distinct intermediate in the transition region. V132A populates this intermediate starting at ~1.5 M GdmCl, while in V75D the transition begins earlier, at around 1.0 M GdmCl (Figure 2.1). Wild-type γDC has been observed to fit well to three-state models (44, 49) but here the two transitions are not distinguishable from one another in a three-state fit. The resulting two-state fit, however, yields an unusually broad transition when compared to the isolated domains and is therefore considered to be unreliable, emphasizing the usefulness of an Ising model in dissecting the energetics of each transition (see below).

All full-length variants, whether containing mutations in the NTD or the CTD, result in destabilization of the first unfolding transition relative to wild-type, with V75D/V132A the most disruptive and V132A the least. Although V132A is a mutation in the CTD, it mostly impacts the first (presumably NTD-unfolding) transition rather than the second, underscoring the significance of the hydrophobic interface to the stability of the NTD in the context of the full-length γDC. In fact, for V75D/V132A, this domain is already in the unfolding transition at 0 M denaturant, indicative of the intrinsic instability of the NTD in this variant. These data are in agreement with similar studies, albeit under different conditions (e.g., temperature) (45).

The Ising model permits quantification of cooperativity between folding modules in a protein and has been successfully applied to linear repeat proteins wherein subsequent repeats interact via a defined interface (55, 61). Given measurements for the equilibrium unfolding transitions of both a full-length protein and its individual domains, this model can relate the extent of unfolding to both the intrinsic ΔG of each domain and the interfacial ΔG upon association of the two domains at their interface.

Therefore, I used a global-fitting approach for a one-dimensional Ising model describing a two-repeat heteropolymer as described previously (62) to analyze folding transitions for full-length crystallins where both isolated domain transitions could be measured (wild type and V132A) (Figure 2.1C).

Wild-type γDC constructs fit well to this 1D-Ising model (RSSR = 4.11×10^{-4}), with intrinsic folding free energies of -4.5 and -8.2 kcal/mol for the NTD and CTD domains, respectively, in good agreement with the two-state models applied to the equilibrium unfolding curves of isolated domains. I found the interface to be contributing -3.8 kcal/mol in wild-type γDC, consistent with the previous estimate (49) (Table 2.1). Thus, the two domains are strongly energetically coupled.

For V132A, however, the equilibrium unfolding transition is not well-fit by the 1D-Ising model, yielding a non-random distribution of fit residuals. Notably, the second unfolding transition (corresponding to unfolding of the CTD) is stabilized in the full-length construct relative to V132A γDC_{CTD} even when the NTD is fully unfolded. Hence, I surmise that the presence of the unfolded NTD must be interacting with the folded CTD in V132A γDC, an interaction which is not captured by the simplest version of the 1D-Ising model. Inclusion of an interaction term permitting the folded V132A CTD to be stabilized by the unfolded NTD as well as by the folded NTD yields a significantly improved fit (RSSR = 5.12×10^{-4}) (55). This interfacial energy was found to be 1.1 kcal/mol, a reduction in interfacial stability of 2.7 kcal/mol as compared to wild-type.

The slope of the second transition in the V75D/V132A double mutant is larger than the slope of this transition in V132A, V75D, or V132A γDC_{CTD} (*m*-value of 3.9 ± 0.3, as compared to 3.2 ± 0.3, 3.2 ± 0.3, and 2.8 ± 0.3, respectively). According to the linear relationship between solvent-exposed surface area and *m*-value (63), this corresponds to an increase in ΔASA of ~200 $Å^2$, which could indicate some degree of difference in the residues involved in the second transition.

Additional mutation to the CTD interface, in dInt-γDC, does not result in further disruption of the first unfolding transition relative to the single mutation V132A, but does result in a decrease of the CTD unfolding transition by 2.5 kcal/mol (Supplemental Figure 2.1A). This suggests that these further mutations to the CTD interface do not result in additional loss of stabilization of the NTD transition, instead only reducing the intrinsic stability of the CTD.

Finally, for technical reasons, I was unable to carry out a full Ising analysis to account for the effect of V75D. The isolated V75D NTD fragment proved refractory to purification and analysis despite multiple attempts. Based on the above data, I would predict that the isolated NTD of V75D would be intrinsically unstable with most of its stability in full-length proteins provided by interfacial interactions, both because V75D is not predicted to abolish any interfacial contacts, and because the total interaction energy in wild-type exceeds the ΔG of the V75D NTD unfolding transition by 1.3 kcal/mol. The low stability of the first transition of the V75D/V132A double mutant (i.e., when stabilizing interactions with the CTD are reduced) also demonstrates the intrinsic instability of the V75D NTD.

		Equilibrium Transition One (NTD)			**Interface**	**Equilibrium Transition Two (CTD)**		
		ΔG°_{NTD}	**C_M**	***m*-value**	**$\Delta G^{\circ}_{interface}$**	**ΔG°_{CTD}**	**C_M**	***m*-value**
Two-State	**γDC$_{NTD}$**	-4.82 ± 0.53	1.43 ± 0.21	3.38 ± 0.34		--	--	--
	γDC$_{CTD}$	--	--	--		-7.83 ± 0.56	3.17 ± 0.33	2.47 ± 0.18
	V132A γDC$_{CTD}$	--	--	--		-6.21 ± 0.37	2.35 ± 0.20	2.64 ± 0.16
	wt γDC	-5.86 ± 0.17	2.98 ± 0.12	1.97 ± 0.06				
Three-State	**V75D γDC**	-3.55 ± 0.76	1.05 ± 0.04	3.38 ± 0.72		-9.90 ± 0.64	3.30 ± 0.02	3.00 ± 0.19
	V132A γDC	-6.36 ± 1.87	1.47 ± 0.07	4.34 ± 1.26		-9.06 ± 0.78	2.81 ± 0.04	3.22 ± 0.27
	V132A/ V75D γDC	0.00 ± 2.27	0.00 ± 0.82	2.78 ± 2.28		-11.20 ± 0.82	2.87 ± 0.01	3.90 ± 0.29
Ising	**Wild-type γDC**	-4.47 ± 0.21	1.42 ± 0.07	3.15 ± 0.15	-3.84 ± 0.22	-8.21 ± 0.36	3.15 ± 0.06	2.61 ± 0.11
	V132A γDC	-4.53 ± 0.27	1.42 ± 0.08	3.19 ± 0.19	-1.14 ± 0.10 [(a)]	-6.66 ± 0.35	2.34 ± 0.07	2.84 ± 0.15

Table 2.1: Equilibrium two-state, three-state, and global Ising fit parameters for all γDC variants. Error reported is standard error of the fit. ΔG (folding) values are in kcal/mol; C_M values are in M GdmCl; *m*-values are in kcal/(mol* M GdmCl). Confidence intervals calculated from bootstrapped parameters for Ising fits can be found in Table 2.2.

[a] The model for V132A includes a stabilizing interaction between the folded CTD and the unfolded NTD equal to that of the stabilizing interaction between the folded CTD and the folded NTD. A more complicated model permitting these two stabilizing interactions to vary independently of one another did not pass an F-test (p = 0.49) and so was not used.

Wild-type γDC			
RSSR = 4.11×10^{-4}	Mean	5% CI	95% CI
ΔG°_{NTD}	-4.48	-4.89	-4.08
ΔG°_{CTD}	-8.24	-8.97	-7.59
$\Delta G^{\circ}_{interface}$	-3.31	-4.03	-2.52
$\Delta G^{\circ}_{interface\ (NTD\ unfolded)}$	0.02	-0.26	0.27
m_{NTD}	3.15	2.87	3.44
m_{CTD}	2.61	2.41	2.85
V132A γDC			
RSSR = 5.12×10^{-4}	Mean	5% CI	95% CI
ΔG°_{NTD}	-4.54	-5.09	-4.04
ΔG°_{CTD}	-6.74	-7.47	-6.11
$\Delta G^{\circ}_{interface}$	-1.21	-1.52	-0.92
$\Delta G^{\circ}_{interface\ (NTD\ unfolded)}$	-1.21	-1.52	-0.92
m_{NTD}	3.20	2.84	3.57
m_{CTD}	2.88	2.61	3.18

Table 2.2: Bootstrapped parameters for Ising fits. Equilibrium two-state, three-state, and global Ising fit parameters for all γDC variants. ΔG° (folding) values are in kcal/mol; *m*-values are in kcal/(mol* M GdmCl). Values are from 3,000 bootstrapping iterations. For V132A, $\Delta G^{\circ}_{interface}$ was held equivalent to $\Delta G^{\circ}_{interface\ (NTD\ unfolded)}$. Abbreviations: RSSR, reduced sum of square residuals; CI, confidence interval.

2.2.2 Continuous-labeling HDX-MS on variants of γD Crystallin

While the above equilibrium unfolding measurements and Ising analyses permit a quantitative determination of the stability of each folding module and the interfacial coupling between them, they are unable to inform about the conformations of any partially unfolded states or to detect the presence of any low-occupancy, high-energy states under native conditions.

Thus, I set out to assess conformational ensembles of variants of γDC on a sub-global level using HDX-MS, for a more complete understanding of the energy landscape. To

this end, I followed the continuous hydrogen-deuterium exchange of wild-type and several variants of γDC in PBS pH 7.0 with 5 mM DTT at 25 ˚C over the course of 15 seconds to 72 hours, both in the presence of denaturant and in its absence (native conditions).

The kinetics of HDX are commonly interpreted as occurring in one of two noted regimes, referred to as EX1 and EX2. When monitored by mass spectrometry, these two kinetic limits may be distinguished by their mass spectra signatures: EX1 kinetics produce two distinct mass envelopes whose relative populations change over the time of exchange, while EX2 results in the gradual increase in m/z over time of a unimodal mass envelope (64). In both extremes, the exchange process is analyzed via the Linderstrøm-Lang model of HDX (52), where an amide can interconvert between an open, exchange-competent state and a closed, exchange-resistant state with rates k_{op} and k_{cl}. In the open state, exchange can occur at a rate k_{int}, the chemical or intrinsic rate of exchange for that amide. When the rate of closing k_{cl} is slow relative to intrinsic rate of exchange k_{int} ($k_{cl} << k_{int}$), the EX1 limit is reached. The exchange rate between the two observed populations is equal to the rate of structural opening k_{op}. Conversely, in the EX2 kinetic exchange limit, the rate of closing k_{cl} must be fast relative to the intrinsic rate of exchange k_{int} ($k_{cl} >> k_{int}$). In this regime, the rate of hydrogen exchange can be related to the free energy of the transition between open and closed forms; specifically, the observed rate constant of exchange can be found as $k_{ex} = k_{op}/k_{cl} * k_{int} = K_{op} * k_{int}$, where K_{op} is the equilibrium constant for opening to the exchange-competent state. Thus, the free energy for the opening transition is $\Delta G_{op} = -RT \ln(k_{int}/k_{ex})$. When the exchange is monitored at the peptide level by mass spectrometry, the rate constants k_{int} per peptide can be estimated based on sequence of the individual peptide (65). Thus, when $\Delta G_{unfolding}$ of a particular structural transition is known, HDX rates per peptide associated with that opening transition can be estimated or predicted. Dynamic or flexible peptides whose opening transitions do not depend on this estimated unfolding reaction may exchange more quickly, but peptides that exchange more slowly in EX2 than this prediction must involve a transition with a greater ΔG. This permits identification of regions of a protein with “extra” stability.

Using aspergillopepsin and pepsin digestion and a two-step quenching method, I obtained between 95-100% peptide coverage of all variants of γDC. This allowed me to interrogate the dynamics of all of γDC (average redundancy, coverage, and back exchange as well as number of replicates for each construct in Table 2.3). In most variants studied, nearly all peptides exhibit a single isotopic envelope distribution whose centroid mass increases as the time of exchange increases, demonstrating EX2 kinetic behavior, with a few notable exceptions.

	V75D	γDC$_{NTD}$	γDC$_{CTD}$	V132A	V132A γDC$_{CTD}$	V75D/ V132A	Wild type	W42R	dInt
Replicates	5	2	2	3	2	2	2	2	1
Back exchange	34.0%	29.7%	28.2%	30.4%	28.6%	25.9%	26.4%	29.2%	25.3%
Coverage	90.6%	96.3%	84.2%	79.6%	75.3%	83.2%	86.1%	89.3%	87.9%
Redundancy	10.6	23.1	4.2	7.5	2.7	5.7	12.1	10.9	9.9
Avg peptide length	11.1	12.4	8.5	11.9	7.9	10.1	11.1	11.0	11.8
Avg # peptides	192.4	176.5	58	132	42.5	120.5	212	210	175
Repeatability	0.31	0.23	0.26	0.22	0.11	0.22	0.28	0.39	

Table 2.3: Average redundancy, coverage, and back exchange as well as number of replicates per each construct studied with HDX-MS. Replicates given as number of technical replicates collected. Redundancy given as average number of peptides per amide. Repeatability given as average standard deviation of deuteration of all peptides across replicates. Average back exchange calculated from angiotensin-II as 100% - #D/(5*90%) (angiotensin-II has five exchangeable protons). Coverage given as percentage of residues contained in at least one peptide. Redundancy given as average number of peptides per amide. Repeatability given as average standard deviation of deuteration of all peptides across replicates.

Uptake data for peptides in the EX2 regime were corrected for back-exchange conditions between different experiments by normalization to exchange of a fiduciary peptide (angiotensin-II) included in all exchange experiments. Corrected data were then fit to a multiexponential model to permit calculation of half-lives by finding the time of half-maximum exchange (by comparison to a fit maximum or to deuteration of an unfolded control). Half-lives found in this manner were used to estimate expected ΔG values for comparison to states identified by equilibrium denaturation.

Time-dependent mass spectra in the EX1 regime were globally fit to a sum of two Gaussian distributions where the heavier peak corresponds to the open state and the lighter peak corresponds to the closed state. Fractional populations at each timepoint were calculated and fit to single-exponential kinetics to obtain rates of opening.

2.2.3 γDC hydrogen-deuterium exchange is slow at the interface

I first sought to monitor V75D's known equilibrium intermediate using HDX-MS. As shown in Figure 2.2B, V75D populates an equilibrium intermediate between 1 and 3 M GdmCl. Past NMR characterization of the V75D γDC intermediate (populated in 4.2 M urea) has shown that this intermediate is comprised of an unfolded NTD and a native-like CTD (50).

Figure 2.2A-C summarize HDX-MS data taken under conditions that populate either the native state (0 M denaturant) and this intermediate (4.2 M urea). When equilibrated in 4.2 M urea for at least three days prior to exchange, peptides in the NTD exhibit maximum exchange by 15 seconds, while CTD peptides exhibit exchange behavior comparable to V75D under native conditions. This is consistent with previous observations and permits comparison to a state with a known structure, hereafter termed the “unfolded-interface intermediate”.

Having characterized this intermediate under denaturing conditions, I then asked whether this same unfolded-interface intermediate is also populated under native conditions. Given that the equilibrium fluorescence studies above indicate that this unfolded-interface intermediate has a stability 3.5 kcal/mol higher than the native state, under EX2 exchange in our experimental conditions, I would expect the peptides in the NTD to be completely exchanged within the first hour. Indeed, Figure 2.2A-C show that under native conditions, almost all the peptides in the NTD peptides of V75D are fully exchanged within 60 minutes.

Notably, however, peptides from the NTD that are part of the interface (residues 43-55 and 71-80) are not completely exchanged after an hour, and in fact do not exhibit full exchange even after 96 hours ($3x10^6$ sec), with an estimated half-life of exchange near 10^4 or 10^5 seconds. This anomalously slow rate corresponds to a predicted $\Delta G_{opening}$ greater than 7 kcal/mol (Figure 2.2D), inconsistent with the unfolded-interface intermediate.

To determine whether this unusual behavior in the interface is unique to the V75D mutation, I monitored exchange of another mutant with a destabilized NTD that causes congenital cataracts, the W42R mutation, which destabilizes the NTD by ~5 kcal/mol (42). Much like V75D, W42R is known to exhibit the γDC fold without major alterations in conformation compared to wild type, though the relative orientation of its two domains changes slightly and there is minor structural readjustment around the introduced arginine. Solution NMR spectroscopy has detected a small fraction of protein with partial unfolding in the NTD and has also showed that deuterium uptake across the NTD is broadly faster in W42R than in wild type (42). I find that, like V75D, W42R exhibits high protection along the interface (Figure 2.2E). Indeed, differences in exchange behavior between W42R and its unfolded-interface intermediate in 1.8 M GdmCl are nearly identical to those of V75D (Figure 2.2B, E). W42R exchange in the NTD is broadly comparable to that in V75D (Supplemental Figure 2.2).

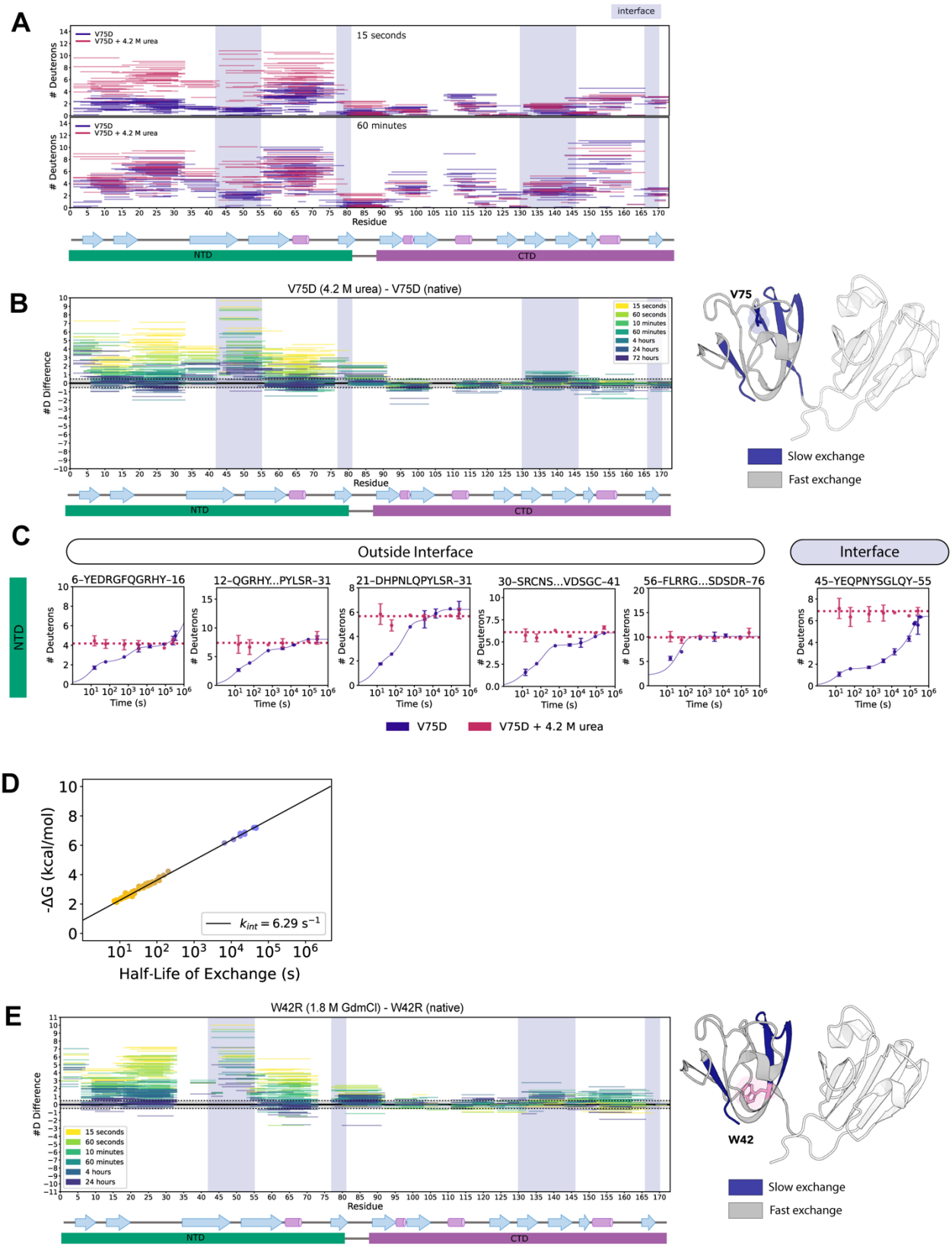

A
interface
V75D
V75D + 4.2 M urea
15 seconds
60 minutes
Deuterons
Residue
NTD
CTD
B
V75D (4.2 M urea) - V75D (native)
#D Difference
15 seconds
60 seconds
10 minutes
60 minutes
4 hours
24 hours
72 hours
V75
Slow exchange
Fast exchange
C
Outside Interface
Interface
6-YEDRGFQGRHY-16
12-QGRHY...PYLSR-31
21-DHPNLQPYLSR-31
30-SRCNS...VDSGC-41
56-FLRRG...SDSDR-76
45-YEQPNYSGLQY-55
Time (s)
V75D
V75D + 4.2 M urea
D
-ΔG (kcal/mol)
k_{int} = 6.29 s^{-1}
Half-Life of Exchange (s)
E
W42R (1.8 M GdmCl) - W42R (native)
W42
Slow exchange
Fast exchange

Figure 2.2: V75D and W42R undergo slow hydrogen-deuterium exchange at the NTD interface. A) Number of deuterons uptaken by V75D under native conditions (blue) and in 4.2 M urea (magenta) at 15 seconds (top) or 60 minutes (bottom). Each line represents an individual peptide spanning the residues indicated on the x-axis, with number of deuterons uptaken indicated on the y-axis. After 60 minutes, exchange profiles of V75D under native conditions mostly match V75D in urea, with notable exceptions at the NTD interface (interfacial regions are highlighted in pale blue). Secondary structure map and domain regions are indicated below the plot. B) Subtractive plot of the data in A (deuteration of V75D in urea– deuteration of V75D in native conditions), for all timepoints listed. A positive difference in deuteration indicates less deuteration (more protection) under native conditions than in urea. Dark blue regions highlighted on the crystal structure are regions of unusually slow exchange. C) Uptake plots for representative peptides in the NTD of V75D, under native conditions (blue) or in 4.2 M urea (magenta). The dotted magenta line is the average of V75D uptake in urea. Error bars represent standard error of technical replicates. The solid blue line is a multiexponential fit to uptake data. D) Graphical illustration of the relationship between half-life of exchange and estimated ΔG for peptides in EX2. Each point is an individual peptide in the NTD of V75D, where a half-life of exchange is found by fitting uptake data to a multiexponential equation and $\Delta G_{peptide}$ is subsequently calculated based on that peptide's amide intrinsic exchange rates. The black line is the average amide intrinsic exchange rate across the entire NTD, $k_{int} = 6.29\ s^{-1}$. Most peptides (yellow) cluster near or under the ΔG of the NTD unfolding transition of V75D, which is 3.5 ± 0.8 kcal/mol. However, some peptides (blue) exchange at a rate corresponding to a significantly higher ΔG. E) Same as B, but depicting data collected from the W42R variant either under native conditions or in 1.8 M GdmCl. Dark blue regions highlighted on the crystal structure are regions of unusually slow exchange.

2.2.4 HDX in isolated domains of γDC

To determine whether slow exchange along the NTD interface relies on interaction with the CTD, I measured continuous hydrogen exchange behavior in the γDC_{NTD} and γDC_{CTD} fragments for comparison to full-length wild-type γDC. I found exchange half-lives consistent with global unfolding of the NTD and CTD at all peptides in γDC_{NTD} and γDC_{CTD} where EX2 behavior could be measured, including along the interface. I compared these exchange behaviors with that of the full-length wild-type γDC (Figure 2.3A, B), along with comparison to V132A γDC equilibrated and unfolded in 4.8 M GdmCl.

As expected, peptides along the interface exchange faster in isolated domains than in full-length γDC (Figure 2.3B). As noted in the equilibrium fluorescence studies above, in the context of the full-length protein, the NTD is stabilized by the interaction energy and therefore has an unfolding free energy of 8.1 kcal/mol, corresponding to a predicted half time for hydrogen exchange of ~ 6 days. Therefore, in the context of the wild-type

protein, I cannot make a determination as to whether or not the interface transition is anomalously slow. Thus, I turned to destabilizing variants to address the role of the interface and the CTD on the hydrogen exchange kinetics noted for V75D and the other variants above.

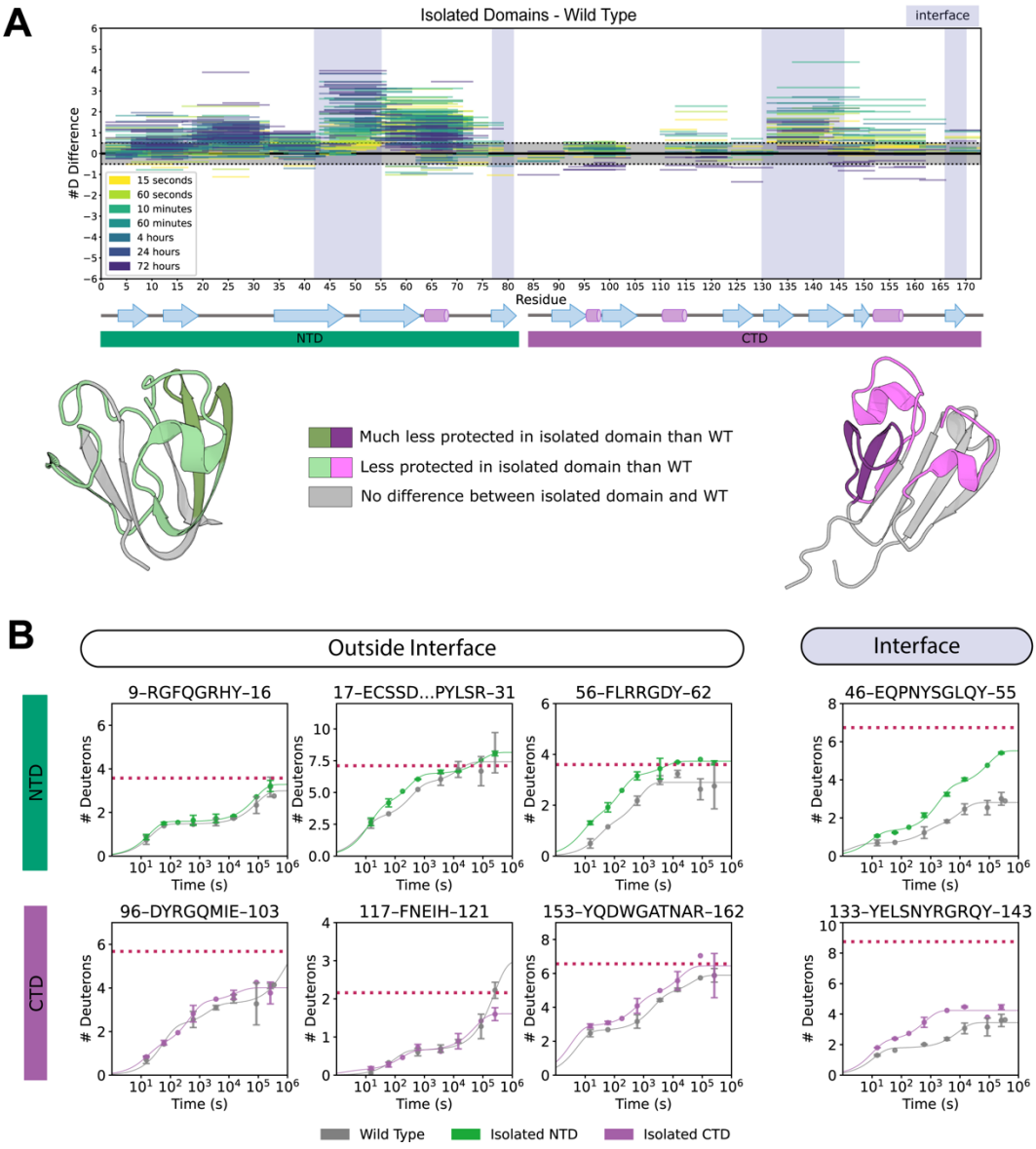


Figure 2.3: Isolated γDC domains lack unusual interface protection. A) Subtractive plot of deuteration of each isolated domain (γDC_{NTD} or γDC_{CTD}) – wild-type γDC. Highlighted regions on each crystal structure (γDC_{NTD}, left; γDC_{CTD}, right) indicate notable differences in deuteration. B) Representative uptake plots for comparison of wild-type and isolated domains. The dotted magenta line is average deuteron uptake for V75D in 4.2 M urea (fully exchanged comparison).

2.2.5 Disruption of the interface in the CTD destabilizes the native state intermediate

To probe the role of interactions with the CTD in contributing to the behavior of the NTD, I destabilized the interface via mutations in the CTD and examined the effect on the hydrogen exchange of peptides in the NTD. If interfacial contacts with the CTD are contributing to the stability of the intermediate populated under native conditions, then I hypothesized that weakening these contacts via the V132A mutation should destabilize that intermediate. I carried out continuous HDX on both V132A and V75D/V132A to measure the impact of CTD-interface mutations in both wild-type and NTD-destabilized contexts.

In the V132A variant, most areas of the NTD (with the exception of those in the slow-exchanging interface) exchange at a rate similar or faster than the same region in the isolated NTD (γDC_{NTD}). Thus, this mutation in the CTD allosterically destabilizes the NTD. Rates of exchange in slow-exchanging interface peptides are greater for V132A than for wild-type γDC, although not as fast as for the isolated γDC_{NTD} (Figure 2.4A, B). This suggests partial, though not complete, loss of tight interfacial interactions. The CTD of V132A is slightly deprotected relative to wild type, especially at interface residues, where it nearly resembles the wild-type γDC_{CTD}. The isolated γDC_{CTD} variant carrying the V132A mutation exhibits minor loss of protection in peptides near the site of mutation, but otherwise closely resembles the wild-type γDC_{CTD} (Supplemental Figure 2.3).

In an attempt to weaken the interface further, I added four more mutations at the interface, all in the CTD, to yield the dInt-γDC construct (V132A, Q143A, L145A, M147A, and V170A) (45, 54), and monitored its stability as above via hydrogen exchange. HDX behavior suggests that these additional mutations in the CTD interface destabilize the CTD rather than further weakening the interface, consistent with denaturant-induced unfolding. Exchange of peptides in the NTD of dInt-γDC is mostly indistinguishable to V132A alone: as in V132A, most of the NTD in this full-length construct resembles the isolated γDC_{NTD}, while exchange at the interface is more rapid than in wild type but slower than in the isolated γDC_{NTD} (Supplemental Figure 2.1). There is destabilization near the sites of mutation in the CTD interface, but no significant changes of note in the NTD interface. Given these results, I did not pursue this variant further and instead used the single V75D mutation to explore the role of the interface in the newly identified intermediate accessible under native conditions.

To probe the effect of a weakened interface on the intermediate identified under native conditions in V75D, I carried out HDX-MS on the V132A/V75D variant and compared the HDX behavior to that of V75D. Compared to V75D, the V132A/V75D variant exhibited decreased protection across the NTD, with most dramatic changes at the interface (Figure 2.4C, D, E, F). In peptides at the interface, V132A/V75D demonstrates bimodal exchange (residues 43-55, 72-80) (Figure 2.4F). This bimodal behavior, however, is not simply a result of EX1 kinetics (as exemplified in peptide 1-7, Figure

2.4F), as the centroid (or average mass) of the lighter peak increases with time. Rather, this behavior indicates the presence of two different populations with a slow rate of interconversion between them. The EX2 hydrogen exchange of the lighter population at the interface of V57D/V132A remains slow (compared to the rest of domain), but it is faster than in V75D (Figure 2.4D). Thus, some degree of destabilization of this intermediate is suggested, though competition with the EX1 exchange pathway makes quantification difficult.

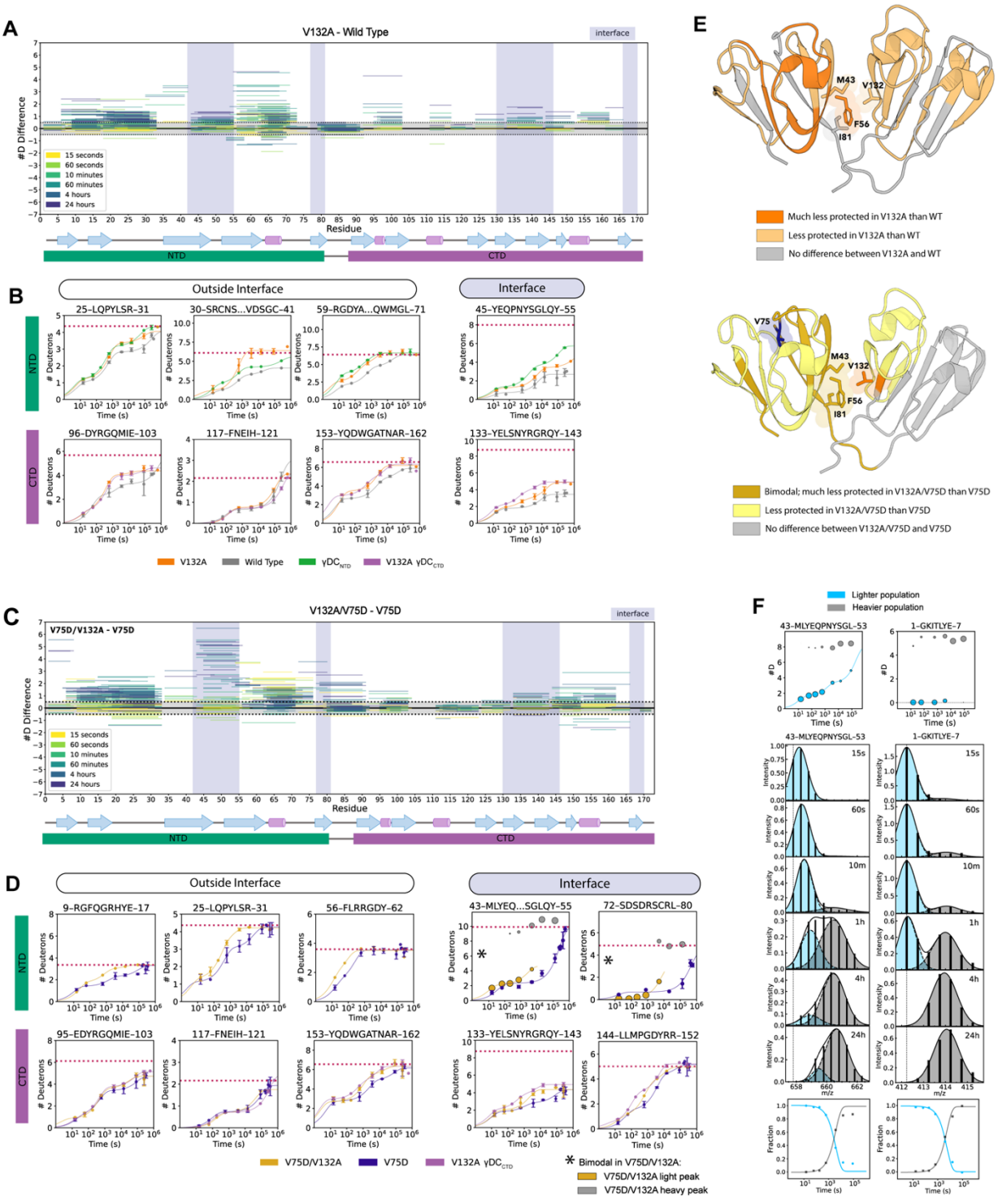


Figure 2.4: Disruption of the CTD interface with the V132A mutation destabilizes the native state intermediate. A) Subtractive plot of deuteration of V132A γDC – wild-type γDC. B) Uptake plots of V132A, wild-type, and isolated domains. The dotted magenta line is average deuteron uptake of V132A in 4.8 M GdmCl (fully exchanged comparison). C) Subtractive plot of deuteration of V132A/V75D γDC – V75D γDC. D)

Uptake plots depicting V75D (dark blue) and V132A (yellow). The dotted magenta line is V132A in 4.8 M GdmCl (fully exchanged comparison). Peptides at the NTD interface in V75D/V132A (marked with an asterisk) demonstrate bimodal behavior and are plotted in yellow and gray to represent the lighter and heavier populations, respectively, where the size of the dot represents the relative population. E) Highlighted regions on the crystal structures indicate notable differences in deuteration (top: V132A vs. wild-type; bottom, V75D/V132A vs. V75D). V132A and its three closest interacting residues in the NTD are highlighted. F) Top: Deuteration and relative population (size of each dot) of the lighter (light blue) and heavier (gray) populations over time. Middle: Representative mass spectra of two peptides in V75D/V132A as time of exchange increases, with overlaid Gaussian fits to indicate two populations (light blue, lighter, less-exchanged population; gray, heavier, more-exchanged population). The first peptide (residues 43-53) is from the interface; the second peptide (residues 1-7) is from the β1 strand. Bottom: Kinetics of conversion between the two populations. Highlighted regions on the crystal structure indicate notable differences in deuteration between V132A/V75D and V75D, with dark mustard corresponding to regions that are bimodal in V132A/V75D.

2.2.6 Peptides with EX1 kinetics report on a correlated opening reaction suggestive of slow interconversion between two states

I analyzed areas of EX1 exchange in order to find opening rates corresponding to the exchange reaction for these sites. In most experiments, I observed only a few peptides that appeared to be in the EX1 regime. For these regions, the observed HDX exchange rate reports on the kinetics of the opening reaction (a transition from an exchange-incompetent, or closed, state to an exchange-competent, or open, state). These regions were observed in two locations: the first beta strand of the NTD, and at the NTD slow-exchanging interface.

The first beta strand of the NTD (residues 1-7, β1) exhibits EX1 behavior in every γDC variant studied (including, surprisingly, γDC_{NTD}), indicating the presence of slow interconversion between the open and closed state involving this strand (Figure 2.5A). Interestingly, neither of β1's hydrogen-bonding partners, β2 and β4 in the NTD, share this EX1 behavior.

I found that the opening rate of β1 increases with destabilization of the NTD, both by mutation and by addition of denaturant. β1 opening is fastest in V75D/V132A ($k = 2.5 \times 10^{-4}$ s^{-1}), which is faster than in V75D ($k = 2.5 \times 10^{-6}$ s^{-1}), which is faster than in γDC_{NTD} ($k = 6.1 \times 10^{-7}$ s^{-1}), which is faster than in wild-type γDC and V132A γDC ($k < 10^{-7}$ s^{-1}, too slow to be measured on the timescale of our experiments) (Figure 2.5B, Table 2.4). When V75D is permitted to equilibrate in increasing amounts of urea prior to exchange (0.6 M - 3.0 M), the natural log of the opening rate of β1 increases linearly (Figure 2.5C). Comparable rates measured by extrapolation from unfolding in high

GdmCl by Mills-Henry *et al.* were ln(k) = -20.6 for unfolding of the NTD in wild type, and ln(k) = -9.6 for global unfolding of γDC_{NTD} (49).

Bimodal exchange at the NTD interface can only be observed upon destabilization of the interface either by mutation (in V75D/V132A) or by equilibration in urea. Upon destabilization by a) V132A mutation or b) denaturant, interface peptides of V75D (residues 43-56) acquire bimodal behavior, involving a slow-exchanging population (blue), and EX1 opening to a fast-exchanging population (gray) (Figure 2.4F, Figure 2.5C).

Under both of these circumstances, rates of EX1 opening for peptides at the NTD interface correlate strongly with rates of EX1 opening at β1. The EX1 opening rate of V75D/V132A at the interface ($k = 2.2 \times 10^{-4}$ s^{-1}) is indistinguishable from its opening rate at β1 (Figure 2.5B), and the linear increase in EX1 opening rates of the V75D interface in urea is indistinguishable from the linear increase in rates of opening at V75D β1 (p = 1.0 by one-way ANCOVA) (Figure 2.5D).

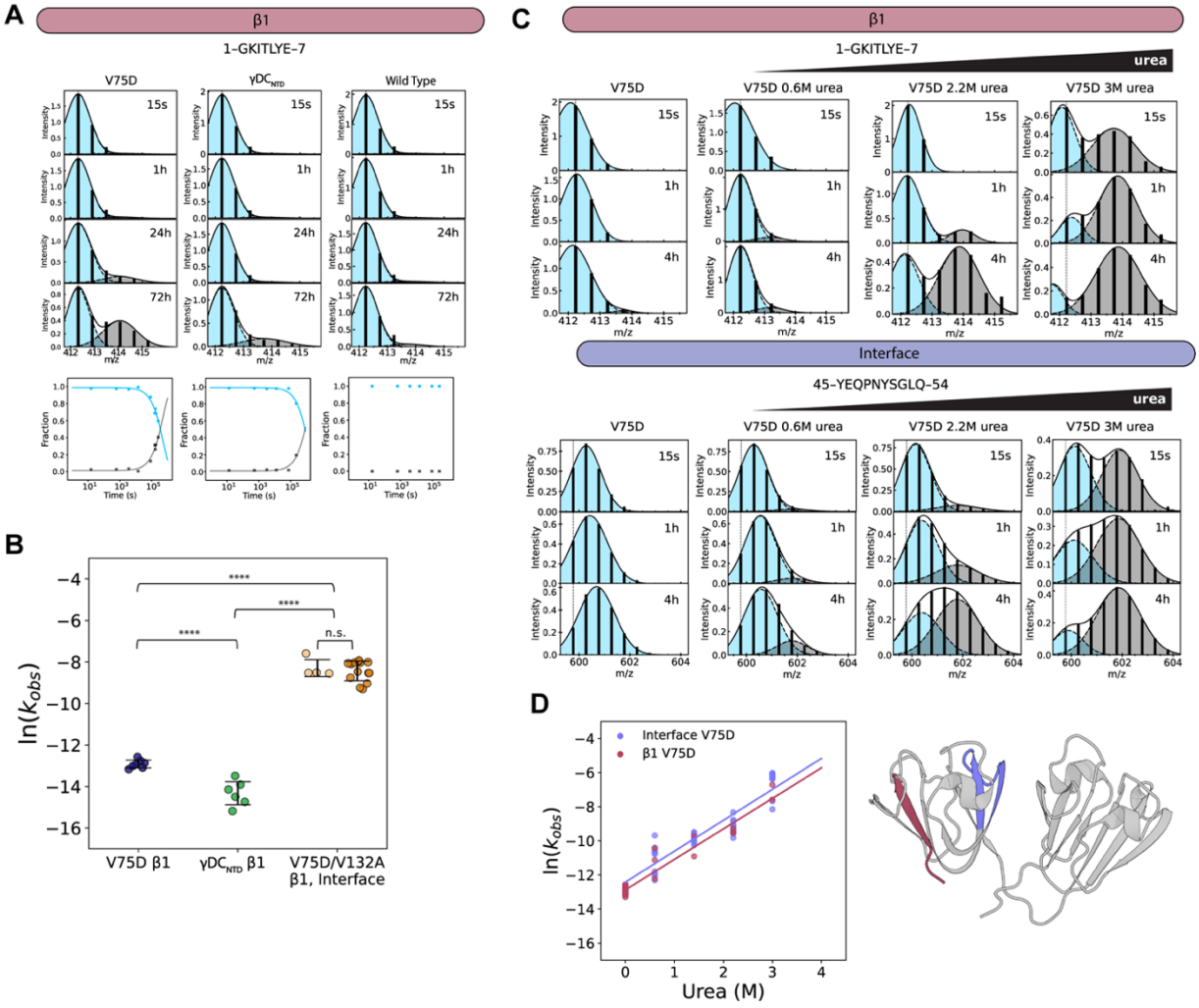


Figure 2.5: Rates of exchange at β1 and the slow interface correlate. A) Top: Representative mass spectra of a β1 peptide in V75D, γDC$_{NTD}$, and wild type as time of exchange increases, with overlaid Gaussian fits to indicate two populations (light blue, lighter, less-exchanged population; gray, heavier, more-exchanged population). Bottom: EX1 kinetics of conversion from light to heavy (wild-type too slow to fit). B) Comparison of rates of EX1 β1 exchange in V75D (dark blue), NTD (green), and V75D/V132A (light orange), along with rate of conversion between slow and fast V75D/V132A rates at the interface (dark orange). Rate values can be found in Table 2.4. Asterisks indicate $p < 0.0001$ by one-way ANOVA and Tukey HSD. C) Example mass spectra of a V75D peptide in β1 and at the interface when equilibrated in increasing amounts of urea. D) Rates of V75D at both β1 (light red) and the NTD interface (slate blue). The linear increase in EX1 opening rates of the V75D interface with respect to urea is indistinguishable from the linear increase in rates of opening at V75D β1 ($p = 1.0$ by one-way ANCOVA). β1 (light red) and the NTD interface (slate blue) are highlighted on the γDC crystal structure.

		ln(k_{op})	# peptides
β1 strand	**V75D γDC**	-12.9 ± 0.19	7
	γDC$_{NTD}$	-14.3 ± 0.55	6
	V75D/V132A γDC	-8.3 ± 0.4	4
	Wild-type γDC	< -15	*
	V132A γDC	< -15	*
Interface	**V75D/V132A γDC**	-8.4 ± 0.46	15

Table 2.4: Apparent EX1 HDX rates of exchange. Rates were measured at two different regions in γDC variants, the β1 strand (residues 1-7) and the interface (residues 43-55), reported as the mean of rates (in s^{-1}) obtained from distinct peptide population transitions individually fit to exponential kinetics. Error reported is standard error of the mean. Peptides were chosen from two (V75D/V132A, γDC$_{NTD}$) or four (V75D γDC) independent exchange experiments. *Wild-type and V132A γDC exchanged too slowly to transition in EX1 on the timescale of our experiments.

2.3 Discussion

The objective of this study was to identify and characterize intermediates on the energy landscape of γD-crystallin. By employing both hydrogen-deuterium exchange/mass spectrometry and traditional chemical denaturation, I aimed to a) discern the structural and energetic characteristics of γD-crystallin's equilibrium intermediate under non-denaturing conditions, and b) discover how modifying or eliminating interfacial interactions between γD-crystallin's two domains can influence those characteristics.

I find that, for two cataract-prone variants of γD-crystallin, the primary equilibrium intermediate populated under native conditions is structurally and energetically distinct from the intermediate populated under chemically denaturing conditions. In addition, these studies highlight the critical role the interdomain interface plays in formation of this new intermediate. Disrupting the interface, either by mutation or by mild denaturation, enables direct observation of both intermediates simultaneously. This newly identified partially folded conformation of γD-crystallin may be significant to aggregation and to cataract formation, underscoring the importance of studying energy landscapes under conditions that emulate physiological environments.

2.3.1 Ising analysis quantifies interdomain interaction energetics

By an Ising analysis of both wild-type and V132A γD-crystallin, I quantified the loss in cooperativity upon mutation of the interface between the NTD and CTD of γDC. In the context of full-length constructs, interface interactions between the two domains of γDC

contribute to stabilization of both NTD and CTD transitions, with the interface contributing more to NTD stability than to CTD stability. This interfacial interaction is crucial for the folding and stability of the NTD – particularly in the V75D variant, in which the NTD is intrinsically destabilized. By a modified 1D-Ising analysis of full-length V132A γDC and the V132A γDC_{CTD} (in conjunction with wild-type γDC_{NTD}), the interfacial energy of V132A is 1.1 kcal/mol, a reduction in the overall strength of the interface by 2.7 kcal/mol. Surprisingly, this interfacial interaction occurs even when the NTD is unfolded. This interaction strength is likely very similar in V75D/V132A, as V75D does not impact the CTD transition.

Fascinatingly, this implies that the specificity of folded-folded contacts between the two domains are lost upon mutation of V132, and that stabilizing interfacial interactions in V132A are just as likely to take place with the unfolded NTD as with the folded NTD. This is not the case for wild type: applying a folded-unfolded interaction term to wild type yields an interaction energy of ~0 kcal/mol, showing that this effect is specific to disruption of the interface (Table 2.1).

This could have structural implications for the equilibrium intermediate of V132A under denaturing conditions: from these data, it's possible to posit that this equilibrium intermediate includes some structure at the interface rather than consisting of a totally unfolded NTD. This is consistent with the observation that V132A retains protection at the NTD interface even when the rest of the NTD is destabilized (Figure 2.4A).

2.3.2 The native-state intermediate is distinct from the denatured intermediate

By HDX examination of sub-global unfolding under native conditions, I identified an equilibrium intermediate of V75D γDC distinct to that populated under denaturing conditions. This newly described intermediate consists of a partially folded NTD and a fully folded CTD, and I term it the "folded-interface intermediate."

Wild-type γDC, with a total stability of ~16 kcal/mol, is too stable for HDX to report on global unfolding within the time frame of my HDX-MS experiments. The same is true of the wild-type γDC_{CTD}, with a global stability of 7.8 kcal/mol. Therefore, any peptide that exchanges fully in the time frame of these experiments is either the result of a partial unfolding event and population of an intermediate, or dynamics in the native structure.

The free energy of the unfolding transition of the V75D NTD to the unfolded-interface intermediate, including dissolution of interfacial interactions with the CTD, is 3.5 kcal/mol. The average intrinsic half-life of exchange of all amides in the γDC NTD under our experimental conditions is calculated to be 0.22 seconds (minimum 0.019 s, maximum 0.67 seconds) (66). This corresponds to an average half-life of approximately 80 seconds in the NTD. Under the supposition that the intermediate occupied by V75D under native conditions has an unfolded NTD, I would therefore expect all peptides in the NTD to reach half-exchange in under 10 minutes, with some variation per peptide

from sequence-based changes in amide intrinsic rates and error in ΔG estimation (Figure 2.2D).

Given the thermodynamic importance of the NTD/CTD interface to NTD stability, especially in a mutant with an otherwise highly destabilized NTD, I expected that interface peptides would be among the most protected regions of the NTD, but that they would either exchange at a rate consistent with the stability of the NTD unfolding transition, or exchange in the EX1 regime if the NTD interface is slow to refold.

However, I found that certain interface peptides are in fact far too highly protected in EX2 for their opening transition to correspond to full unfolding of the NTD, as would be the case in the unfolded-interface intermediate. The slow rate of exchange at residues 43-55 and 73-80 is inconsistent with the unfolding transition of the V75D NTD to the unfolded-interface intermediate (Figure 2.2D). This implies that V75D populates an alternative intermediate containing some degree of structure at the interface. This effect is not limited to the V75D mutation, as W42R displays the same behavior.

Furthermore, rates of exchange of peptides in each isolated domain are consistent with the stability of the isolated domains as measured by GdmCl equilibrium denaturation, demonstrating that a) unusually slow interface exchange only takes place in full-length γDC, and b) slow exchange behavior along the NTD interface is a consequence of interaction with the CTD, rather than being a consequence of some native misfolding or interaction event taking place within the NTD itself.

2.3.3 Mutation and denaturant skew relative population of the two intermediates

The domain interface residues are well-known to be critical to γDC's stability and unfolding rate (54). Here, I find that they are also involved in the folded-interface equilibrium intermediate. I also demonstrate the ability to alter which intermediate is populated via mutation at the interface, and I demonstrate the ability to observe two distinct intermediate populations at once.

Mutation of a single residue in the CTD impacts the population of the folded-interface intermediate. Of the five CTD interface mutations studied here, V132A is sufficient to make most of the NTD display isolated domain-like exchange behavior. This signifies that the stabilization conferred allosterically throughout the NTD as a result of the interface is disrupted by V132A, causing most of the V132A NTD to adopt single domain-like behavior. However, the slow interface is more difficult to disrupt, and V132A only partially impacts it.

β1 is in relatively fast EX1 exchange in V75D/V132A, but in slow EX1 exchange in V75D and in all other constructs, including γDC_{NTD}. In an isolated peptide of residues 1-10 of γDC, the β1 sequence exchanges rapidly (full exchange in < 15 s), indicating that the peptide itself has no ability to form intrinsic structure (data not shown). Because this EX1 behavior occurs in isolated γDC_{NTD}, which does not populate the folded-interface intermediate, it must signify a transition to the unfolded-interface intermediate. This is

consistent with the finding that the NTD has slow global unfolding as observed previously (49). I therefore propose that β1 is a reporter on the rate of EX1 opening to the unfolded-interface intermediate.

The interface undergoes slow EX2 exchange in most constructs, including V75D. However, in V75D/V132A, bimodal exchange is observed at the interface (Figure 2.4F). One population at the interface undergoes slow EX2 exchange, and the other population undergoes faster, likely EX1 exchange. This fast population cannot correspond to global unfolding in EX2, and there are no other obvious candidates as an alternative EX2 transition that would not be in fast exchange with the folded-interface intermediate. Therefore, it is most likely that while the first, slow EX2 population is following usual interface kinetics associated with the folded-interface intermediate, the second, faster-exchanging population indicates EX1 transition to the unfolded-interface intermediate, and therefore that V132A destabilizes V75D such that both intermediates are significantly populated.

The interface undergoes similar bimodal exchange upon destabilization of the V75D interface by urea (Figure 2.5C). In all cases, EX1 rates at the interface are equivalent to those observed at β1 (Figure 2.5D), demonstrating that β1 and the interface are likely undergoing EX1 exchange to the same state. However, in most constructs, any observation of EX1 exchange to the unfolded-interface intermediate is precluded by interference of slow EX2 exchange to the folded-interface intermediate.

2.4 Conclusions

Many proteins associated with aggregation-related diseases have a common feature: the ability to populate a non-native or partially folded conformation which is more aggregation-prone than the native state (12). Characterizing the means by which environmental and sequence factors influence this mis- or unfolding is critical to understanding these diseases.

Here, I show that the dominant intermediate accessible to variants of γDC under native conditions involves only partial unfolding of the NTD, rather than the full unfolding that is well-known to occur under mildly denaturing conditions. This intermediate is hidden to structural and biochemical analysis that requires the use of denaturing conditions, and as such, had not been characterized before.

The precise involvement of β1 with this hidden intermediate is unknown, but it is interesting to note that an aggregation swapping mechanism based on the exchange of an N-terminal β-stand to form an intertwined dimer has been identified in nitrollin, a protein which forms a single domain βγ-crystallin fold (67). Moreover, several congenital cataract point mutations tend to cluster near the N-terminal β-strand (including L5S (41), F9S (68), and P23S/T (69)), which may also be a signifier of this strand's relevance in the mechanism of aggregation (70). The N-terminal β-strands have also been

implicated as candidates for involvement in γDC aggregation precursors both by oxidation-induced aggregation in W42Q (71, 72) and by single-molecule domain swapping in wild-type γDC (73).

My results allow me to propose a model for V75D γDC's energy landscape (Figure 2.6). The dominant intermediate has both a free energy closer to that of the native state and a lower kinetic barrier than the intermediate observed under denaturing conditions. The relative population of each intermediate can be changed with denaturant or by mutation of the interface. By lowering the kinetic barrier toward interface unfolding, I observe simultaneous population of both intermediates.

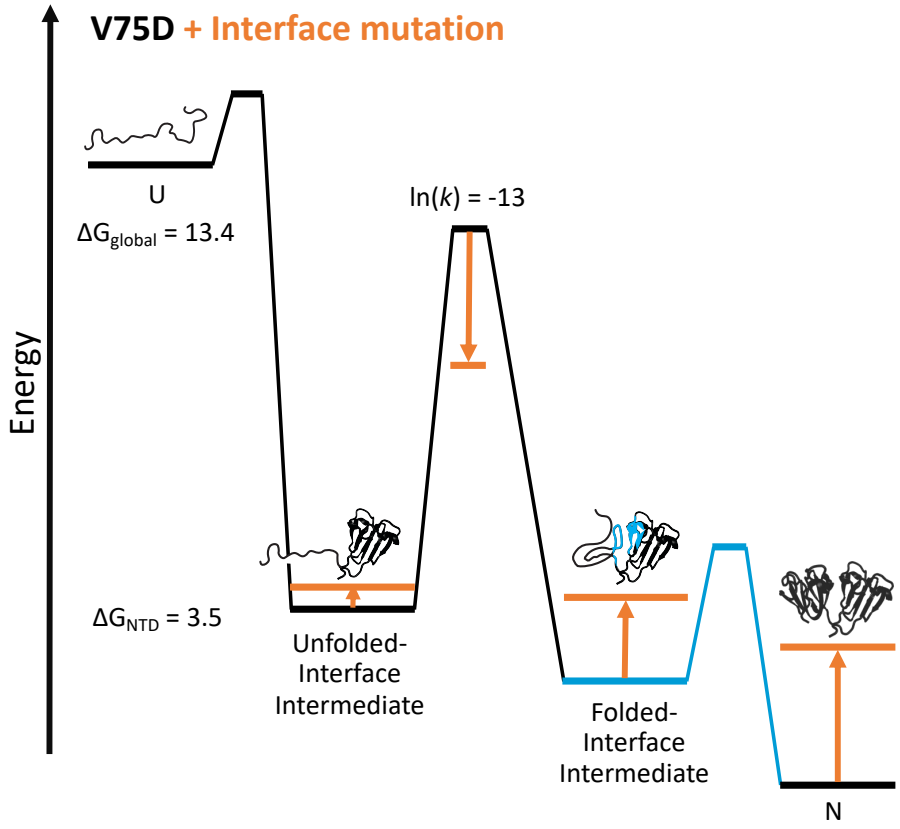


Figure 2.6: Model of the energy landscape of V75D γDC. In a moderately denaturing context, NTD-destabilized γDC (V75D or W42R) populates the unfolded-interface intermediate. Under native conditions, the most-populated intermediate is the folded-interface intermediate, with a high kinetic barrier restricting access to the unfolded-interface intermediate. The V132A mutation's impacts on the energy landscape are illustrated in orange (some effects are estimated). Destabilization of the interface either by mutation or by mild denaturant lowers the kinetic barrier to population of the unfolded-interface intermediate, permitting observation of both intermediates at once. ΔG (unfolding) values are in kcal/mol.

This work does not establish whether the identified hidden intermediate is aggregation-prone, nor does it directly identify a link between interface stability and aggregation. However, structurally distinct equilibrium and folding intermediates have been shown to be crucial precursors in aggregation pathways for various diseases (74). Moreover, partially folded species may be relevant to the types of conformers encountered and bound by α-crystallins (75).

This conformation may be important along the pathway to aggregation and cataract formation. *In vitro* aggregation pathways have been previously described for γDC upon rapid refolding from high denaturant concentrations (46), but it is unlikely that these proteins may encounter such conditions *in vivo*. Such refolding intermediates could be relevant in the context of aggregation upon folding, but in the context of aggregation from native conditions it is important to consider excursions from the native side of the folding barrier.

The observation of differential stabilities for γDC's two domains has led to study of domain swapping as a possible mechanism for γDC aggregation (48). Contrary to the simplest version of the domain-swapping model, in which the entire NTD and CTD may be exchanged between monomers, the partially unfolded intermediate described here retains substantial structure at the NTD-CTD interface. Increased conformational flexibility across the NTD caused by a destabilizing mutation, combined with the local stabilizing effect of the NTD-CTD interface, results in a partially unfolded NTD, which exposes a surface that is normally buried both in the full-length structure and in folded γDC_{NTD}. Prevailing models for cataract formation invoke off-pathway folding, and it is essential to analyze these potential off-pathway intermediates under conditions that emulate physiological circumstances.

2.5 Materials and Methods

2.5.1 Expression and purification of γD-crystallin variants

Mutations in the gene for wild-type γD-crystallin (from (76)) were created via site-directed mutagenesis using primers encoding the proper base pair changes (IDT). Mutations were confirmed by sequencing the region of interest (Quintara).

Wild-type and all mutant γD-crystallins except for the V75D/V132A double mutant were purified as described previously (50). Briefly, for each construct, an *Escherichia coli* BL21 (DE3) colony bearing a pET14b vector encoding the protein of interest was used to inoculate 50 mL of Luria-Bertani (LB) medium supplemented with 100 mg/mL ampicillin. Following overnight growth at 37˚C in a rotary shaker, this culture was used to inoculate 2 L of fresh LB medium containing 100 mg/mL ampicillin at 37˚C in a rotary shaker to an OD600 of ~0.6 and induced with 1 mM isopropyl b-D-1-thiogalactopyranoside for 3 h at 37˚ C. Cells were pelleted and resuspended in ~30 mL of buffer Q (50 mM Tris pH 8.0, 1 mM EDTA, 1 mM dithiothreitol (DTT)) supplemented

with 1× Halt protease inhibitor cocktail (Thermo) and benzonase (Novagen), and stored at 80˚C prior to thawing for protein purification.

Thawed pellets were lysed by sonication, and the lysate was clarified by centrifugation at 15,000 x g for 30 min. The supernatant was filtered with a 0.2 um vacuum filter and passed over a 5 mL HiTrap Q HP anion exchange column (GE Healthcare Life Sciences, Piscataway, NJ) equilibrated in buffer Q. γDC was collected in the flow-through and diluted 1:1 with buffer S (25 mM MES pH 6.0, 1 mM EDTA, 1 mM DTT, and 2% v/v glycerol), after which the pH was adjusted to 6.0 via dropwise addition of HCl. This eluate was filtered with a 0.2 um vacuum filter and loaded onto a HiPrep SP XL 16/10 cation exchange column (Cytiva) equilibrated in buffer S. Bound proteins were eluted using a 0–75% gradient of buffer S plus 1 M NaCl. The crystallin-containing elution peak (as assessed by SDS-PAGE) was passed over either a S75 16/60 size-exclusion column (GE) or S75i 10/300 size-exclusion column (GE) pre-equilibrated in PBS pH 7.0, 5 mM DTT.

To avoid purification of truncated V75D/V132A γDC expression products lacking the NTD, His_6-MBP with a C-terminal tobacco etch virus (TEV) proteolysis site fused to V75D/V132A γD-crystallin (pSV040) was cloned via Gibson assembly and grown and expressed as above. Clarified lysate was allowed to batch bind to HisPur Ni^{2+}-NTA resin (Thermo) washed with 50 mM Tris pH 7.5, 150 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, and eluted with 50 mM Tris pH 7.5, 150 mM NaCl, 500 mM imidazole and 0.5 mM TCEP. Eluate was concentrated in an Amicon spin concentrator (Millipore) and loaded onto a Superdex200 16/60 size-exclusion column (GE) equilibrated in 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP. Peaks corresponding to the MBP-crystallin fusion were collected and digested overnight at 4 ˚C with TEV protease with an N-terminal 6× histidine tag, yielding a final product with no scar. The His-tagged TEV protease and the His-MBP scaffold were removed via a subtractive Ni^{2+}-NTA affinity step using HisPur Ni^{2+}-NTA resin pre-equilibrated with 50 mM Tris pH 7.5, 150 mM NaCl, 20 mM imidazole, 0.5 mM TCEP. Flow through was then concentrated and loaded onto an S75i 10/300 size-exclusion column (GE) pre-equilibrated in PBS pH 7.0, 5 mM DTT.

Peaks corresponding to the correct crystallin constructs were verified by MS, quantified by UV-Vis absorption at 280 nm, and flash-frozen for storage at -80˚C.

2.5.2 Determination of global stability by intrinsic tryptophan fluorescence

For each construct, two 5 μM protein stocks were prepared: a no denaturant protein stock and a high GdmCl (5.6 M) protein stock, both in PBS pH 7.0 (Sigma-Aldrich P4417), 5 mM DTT. Samples with a range of GdmCl concentrations were prepared by combination of the two stocks and allowed to equilibrate at room temperature for at least 24 hours for all constructs except wild-type, which was equilibrated for 172 hours. Measurements were then performed at 25 °C using a PTI Quantamaster Fluorometer (Horiba). An excitation wavelength of 280 nm was used to excite tryptophan residues,

and emission spectra were recorded from 310 to 390 nm (0.7 s/nm). Samples were recovered from the cuvette after each measurement and the exact GdmCl concentration was determined by taking the refractive index. Signal was reported as a ratio of signal at 360 nm to signal at 320 nm.

Signal ratios per concentration GdmCl from each variant were fit to a two-state (equation 1) or three-state (equation 2) folding model using Python's LMFIT module (77), which allowed determination of the transition midpoint, $\Delta G_{unfolding}$ and *m*-value for each transition.

Two-state fit (59):

$$y_{obs} = \frac{F_{intercept} + F_{slope}x + \left(U_{slope} + U_{intercept}x\right)e^{m(x-Cm)/RT}}{1 + e^{m(x-Cm)/RT}} \quad (1)$$

Parameter definitions are as follows: $F_{intercept}$, folded intercept; F_{slope}, folded baseline slope; $U_{intercept}$, unfolded intercept; U_{slope}, unfolded baseline slope; *m*, *m*-value; C_m, transition midpoint. R is the gas constant and T is temperature. Baseline slopes were permitted to vary.

Three-state fit (60):

$$y_{obs} = \frac{F_{intercept} + F_{slope}x + \left(I_{slope} + I_{intercept}x\right)e^{m_1(x-Cm1)/RT} + \left(U_{slope} + U_{intercept}x\right)e^{m_1(x-Cm1)/RT}e^{m_2(x-Cm2)/RT}}{1 + e^{m_1(x-Cm1)/RT} + e^{m_1(x-Cm1)/RT}e^{m_2(x-Cm2)/RT}} \quad (2)$$

Parameter definitions are as follows: $F_{intercept}$, folded intercept; F_{slope}, folded baseline slope; $I_{intercept}$, intermediate unfolded intercept; I_{slope}, intermediate baseline slope (held at zero for all constructs but V132A); $U_{intercept}$, unfolded intercept; U_{slope}, unfolded baseline slope; m_1, transition one *m*-value; C_{m1}, transition one midpoint; m_2, transition two *m*-value; C_{m2}, transition two midpoint. R is the gas constant and T is temperature.

To globally fit unfolding transitions of γDC variants, I generated fitting equations using a one-dimensional Ising model as described previously (55, 78). Modified Python code from Barrick et. al. (78) was used to generate and fit partition functions describing the fraction of a given folded state as a function of denaturant using separate equilibrium constants for folding of individual domains and for the interface coupling. The equilibrium constant κ for an individual domain *i* as a function of GdmCl is $\kappa_i = e^{-(\Delta G_i + m_i * x)/(RT)}$, and the equilibrium constant τ for the interface coupling between the NTD and the CTD is $\tau_{N,C} = e^{-\Delta G_{N,C}/(RT)}$ (Table 2.5).

		1-D Ising Model	Modified 1-D Ising Model
Isolated domain model (two-state)	Partition function	$\rho = 1 + \kappa_i$	$\rho = 1 + \kappa_i$
	Fraction folded	$\frac{\kappa_i}{\rho}$	$\frac{\kappa_i}{\rho}$
Full construct model (four-state)	Partition function	$\rho = 1 + \kappa_N + \kappa_C + \kappa_N\kappa_C\tau_{N,C}$	$\rho = 1 + \kappa_N + \omega_C\kappa_C + \kappa_N\kappa_C\tau_{N,C}$
	Fraction unfolded	$1 - \frac{\kappa_N\kappa_C\tau_{N,C} + \kappa_C/2 + \kappa_N/2}{\rho}$	$\frac{1}{\rho}$
	Fraction partially folded	--	$\frac{\omega_C\kappa_C + \kappa_N}{\rho}$
	Fraction fully folded	$\frac{\kappa_N\kappa_C\tau_{N,C} + \kappa_C/2 + \kappa_N/2}{\rho}$	$\frac{\kappa_N\kappa_C\tau_{N,C}}{\rho}$

Table 2.5: Details of Ising partition functions and fraction folded expressions used for global 1-D Ising analysis of wild-type and V132A γDC. The equilibrium constant κ for an individual domain *i* as a function of GdmCl is $\kappa_i = e^{-(\Delta G_i + m_i * x)/(RT)}$, and the equilibrium constant τ for the interface coupling between the NTD and the CTD is $\tau_{N,C} = e^{-\Delta G_{N,C}/(RT)}$. The equilibrium constant ω describing coupling between the unfolded NTD and the folded CTD is $\omega_C = e^{-\Delta G_{unfoldedN,C}/(RT)}$.

The 1-D Ising partition function used to fit wild-type data contains terms for four states: unfolded, only NTD folded, only CTD folded, and both NTD and CTD folded and interacting. I incorporate native and denatured baselines (locally fit) by multiplying baselines by fraction-folded expressions.

To fit V132A data, I modified the 1-D Ising partition function to permit coupling between the CTD and the unfolded NTD. It still contains terms for four states—unfolded, only NTD folded, only CTD folded and interacting with the unfolded NTD, both NTD and CTD folded and interacting—but an additional equilibrium constant is included to describe coupling between the unfolded NTD and the folded CTD: $\omega_C = e^{-\Delta G_{unfoldedN,C}/(RT)}$. In order to include an intermediate baseline term, I separated the fraction-folded expression into three terms which are multiplied by their respective baseline parameters (locally fit).

2.5.3 HDX-MS continuous exchange

Deuterated buffers were prepared by lyophilizing PBS pH 7.0 containing 5 mM DTT and resuspending it in D_2O (Sigma-Aldrich 151882). All urea- and GdmCl-containing buffers were lyophilized and deuterated a total of three times to ensure total deuteration of the denaturant. Protein samples to be exchanged in the presence of denaturant were equilibrated by incubation in the requisite concentration of denaturant at 25 ˚C in PBS pH 7.0, 5 mM DTT for at least 24 hours. The peptide angiotensin-II (sequence DRVYIHPF, Thermo Scientific) was included in all samples at a concentration of 0.25 µg/mL as a fiduciary to correct for potential variability in back exchange between different buffers and experimental conditions.

To initiate continuous-labeling, samples were diluted tenfold into temperature-equilibrated, deuterated PBS buffer for a final γDC concentration of 7.5 µM. Samples were quenched at 15 s, 60 s, 10 min, 1 hr, 4 hr, 24 hr, and 72 hr by mixing 6 µL of the partially exchanged protein with 24 µL of quench buffer 1 (8.6 M urea, 500 mM TCEP pH 2.2) on ice. Extra time points were collected in some experiments for additional resolution. Quenching samples were incubated on ice for 1 minute to allow for partial unfolding to assist with proteolytic degradation and then were flash frozen in liquid nitrogen and stored at –80 °C.

Samples were thawed by resuspension with 50 µL of ice-cold quench buffer 2 (0.75 M glycine, 50 mM TCEP, pH 2.2) to reduce denaturant concentration prior to proteolytic digestion and were then immediately injected into a cooled valve system (Trajan LEAP valve box) connected to a Thermo Ultimate 3000 LC and Q-Exactive Orbitrap MS using a 250 µL sample loop. Inline digestion was performed with firstly aspergillopepsin (Sigma-Aldrich P2143) and secondly porcine pepsin (Sigma-Aldrich P6887) conjugated to beads (Thermo Scientific POROS 20 Al aldehyde activated resin 1602906) and packed into homemade protease columns. After digestion, peptides were desalted over a hand-packed trap column (Thermo Scientific POROS R2 reversed-phase resin 1112906, 1 mm ID × 2 cm, IDEX C-128). Digestion and desalting took place over 6 minutes: 4 minutes at a flow rate of 50 µL/min, and 2 minutes at a flow rate of 300 µL/min.

Acetonitrile, formic acid, and MS-grade water (Fisher Optima LC/MS) were used to prepare mobile phase solvents A (0.1% formic acid) and B (100% acetonitrile, 0.1% formic acid). Peptides were separated by a linear gradient from 5–40% solvent B over a C8 analytical column (Thermo Scientific BioBasic-8 5-µm particle size 0.5 mm ID × 50 mm 72205-050565) with a guard column attached to the inlet (Thermo Fisher Scientific) for 12 minutes, followed by 40–90% solvent B over 30 seconds. Following peptide elution, analytical and trap columns were subjected to a sawtooth wash and subsequently equilibrated at 5% solvent B prior to the next injection. Protease columns were washed with two (for deuterated samples) or three (for undeuterated samples) injections of 100 µL 1.6 M GdmCl, 0.1% formic acid (passed through a 0.2µm filter). Peptides were eluted directly into a Q Exactive Orbitrap Mass Spectrometer operating in

positive mode (resolution 140,000, AGC target 3×10^6, maximum IT 200 ms, scan range 300–1,500 *m/z*). A tandem mass spectrometry experiment was performed for undeuterated samples of each crystallin construct on every day that data for that construct was collected, in order to generate corresponding peptide lists and retention times (full MS settings the same as above, dd-MS^2 settings as follows: resolution 17,500, AGC target 2×10^5, 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). LC and MS methods were run using Xcalibur 4.1 (Thermo Scientific).

2.5.4 HDX-MS data analysis

Peptides were identified using Byonic software (Protein Metrics) with the yD-crystallin sequence containing the corresponding mutations and the sequence of angiotensin-II used as the search library. Sample digestion parameters were set to non-specific. Precursor mass tolerance and fragment mass tolerance were set to 6 and 40 ppm, respectively.

Peptide lists were imported into HDExaminer3 (Sierra Analytics) along with the deuterated and undeuterated sample mass spectra for analysis. Isotopic distributions were fit and checked manually using HDExaminer3; the monoisotopic mass for each peptide was subtracted from the mass centroid and extracted. Bimodal isotopic distributions were exported from HDExaminer3 for further analysis. Downstream quantitative analysis was performed using Python scripts in Jupyter notebooks.

Bimodal peptide mass spectra for all timepoints were globally fit to a sum of two Gaussian distributions. where the heavier peak corresponds to the open state and the lighter peak corresponds to the closed state. The center and width of each Gaussian was allowed to vary by only 0.2 Da in cases where EX1 behavior was seen. The center and width of the Gaussians were allowed to vary as needed in cases where a distribution centroid was seen to be migrating. Fractional populations at each timepoint were calculated as the area under each Gaussian and fit to a single-exponential kinetics (equation 3) using the LMFIT package to obtain rates of opening (77).

$$y = h - A\mathrm{e}^{(-k*t)} \tag{3}$$

Parameter definitions are as follows: A, amplitude; h, height; k, rate (s^{-1}).

Uptake data in the EX2 regime were corrected for conditions between different buffers by normalization to exchange of a fiduciary peptide included in all exchange experiments, angiotensin-II (equation 4) (79).

$$D_{corr}(t) = \frac{m(t) - m_{0\%}}{A(t) - A_{0\%}} \tag{4}$$

Parameter definitions are as follows: m(t), peptide centroid mass at a given time; $m_{0\%}$, non-deuterated peptide centroid mass; A(t), angiotensin-II centroid mass at a given time

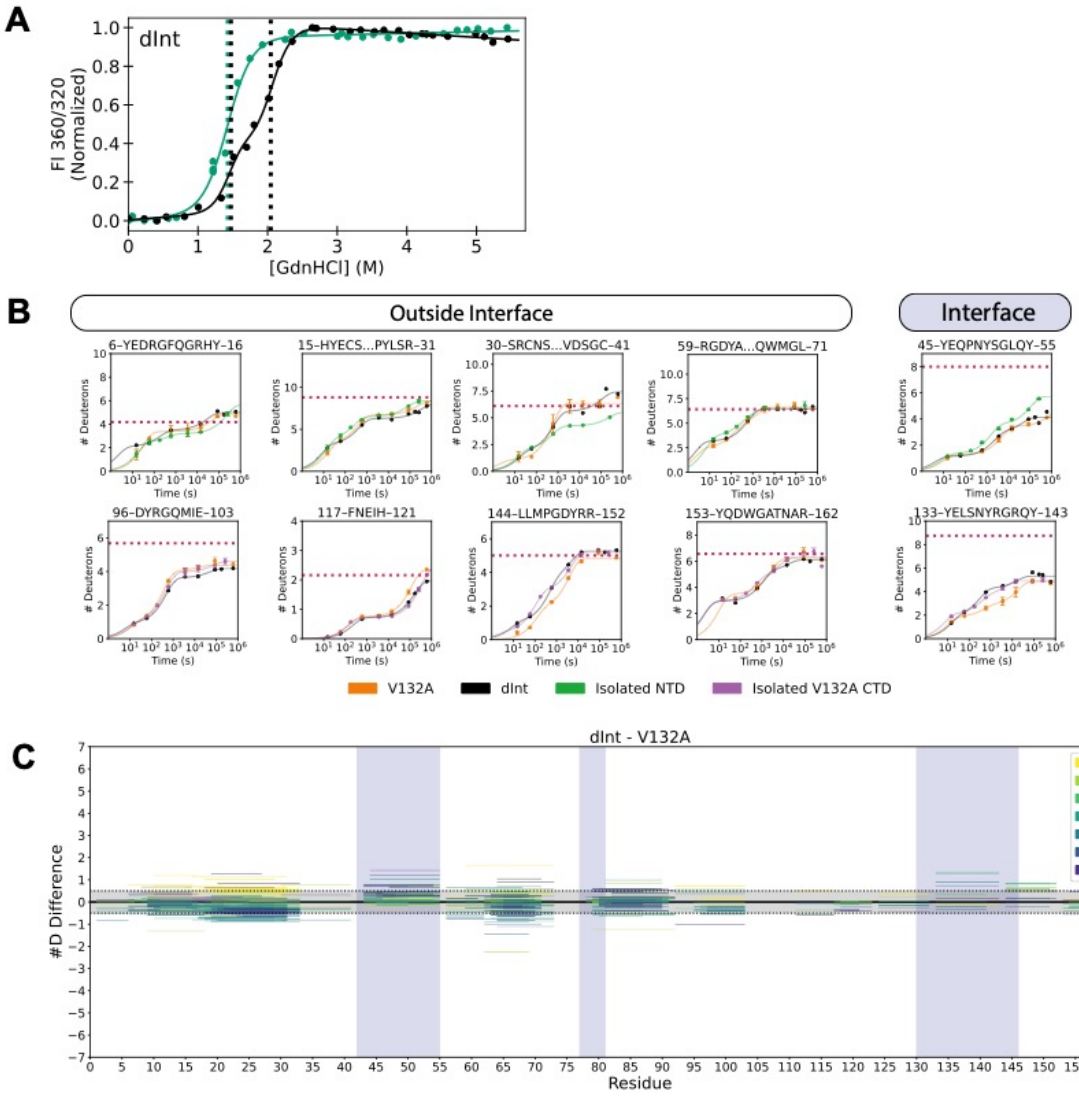
(maximally labeled before 15 seconds); $A_{0\%}$, non-deuterated angiotensin-II centroid mass.

Corrected data were then fit to a multiexponential model (equation 5) using LMFIT (77). For each peptide, either one, two, or three multiexponential terms were used (terms for A, A and B, or A, B and C); inclusion of additional terms in the multiexponential fit was judged automatically on a per peptide basis based on comparison of the Akaike information criteria (AIC) for each model. Each additional term was only included if it decreased the AIC.

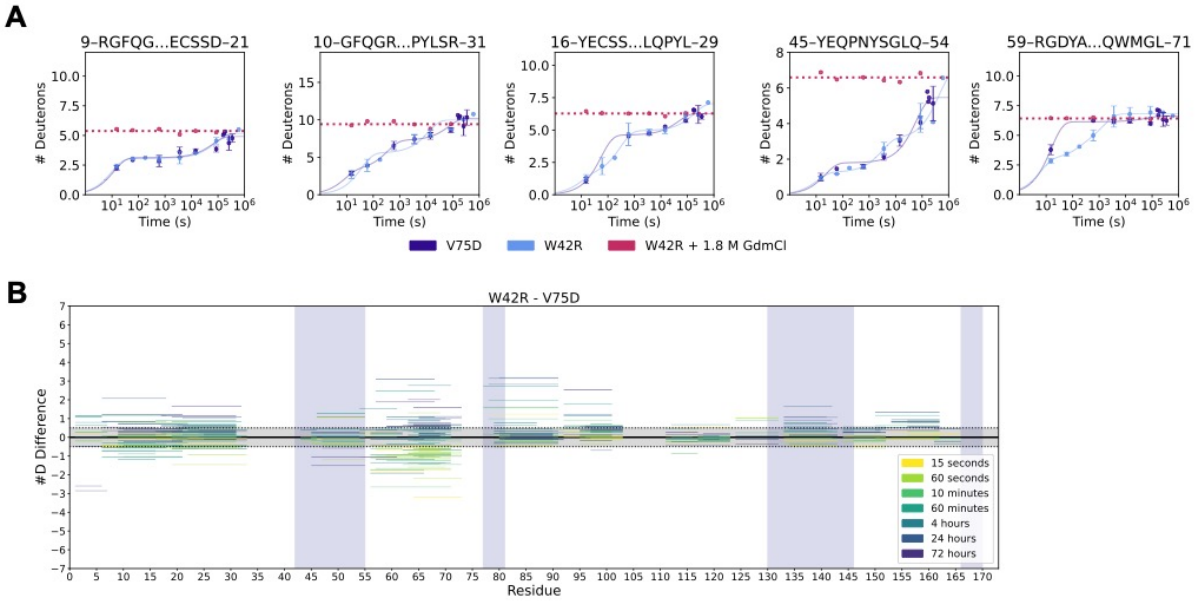
$$D(t) = maxP - Ae^{(-k_A * t)} - Be^{(-k_B * t)} - Ce^{(-k_C * t)} - NE \quad (5)$$

Parameter definitions are as follows: D(t), deuteration; maxP, maximum theoretical exchangeable protons (length of the peptide – 2 – number of prolines); A, B, C: number of fast-, medium-, and slow-exchanging protons; kA, kB, kC: rates of exchange for fast-, medium-, and slow- exchanging protons; NE: number of non-exchanging protons.

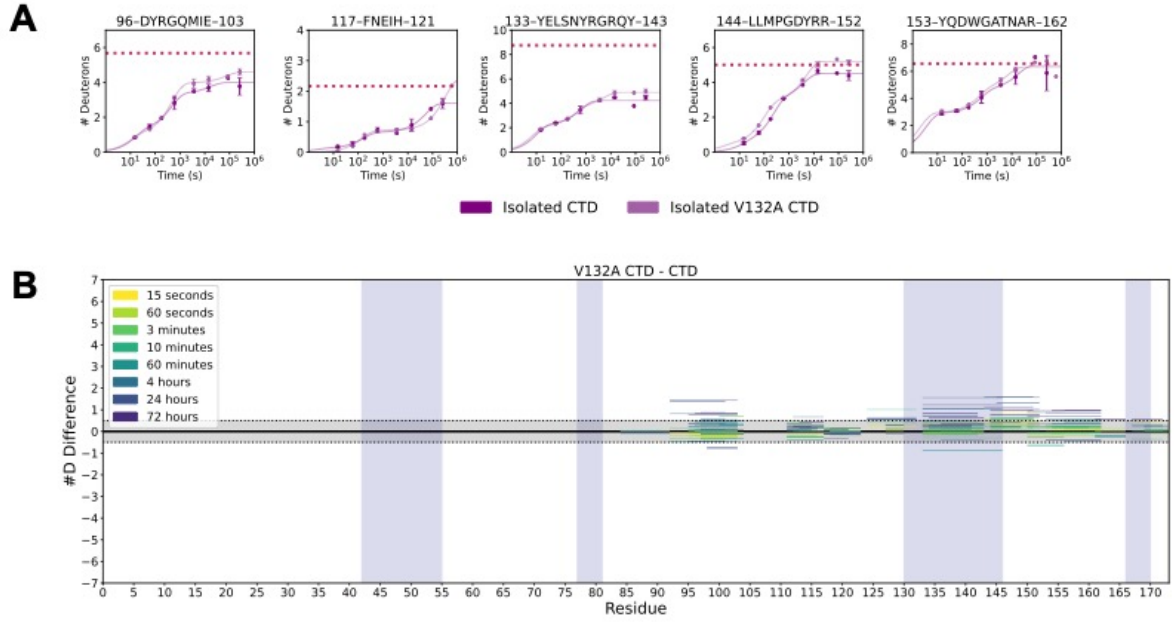
This permitted calculation of half-lives by finding the fit's intersection at half-maximum exchange, as determined by a corresponding fully exchanged control (or by comparison to the fit's calculated maximum exchange). Half-lives found in this manner were used to estimate expected ΔG values for comparison to states identified by equilibrium denaturation melts by using intrinsic rates calculated using *SPHERE* at www.fccc.edu/research/labs/roder (66).



Supplemental Figure 2.1: A) Equilibrium denaturation of dInt γDC (green, γDC_{NTD}; black, dInt γDC). B) Uptake plots comparing dInt γDC, V132A γDC, γDC_{NTD}, and V132A γDC_{CTD}. The dotted magenta line is average deuteron uptake of V132A in 4.8 M GdmCl (fully exchanged comparison). Exchange of peptides in the NTD of dInt-γDC is mostly indistinguishable to V132A alone: as in V132A, most of the NTD in this full-length construct resembles the isolated γDC_{NTD}, while exchange at the interface is more rapid than in wild type but slower than in the isolated γDC_{NTD}. C) Subtractive deuteration plot of dInt γDC and V132A γDC.



Supplemental Figure 2.2: A) Uptake plots comparing W42R, V75D, and W42R + 1.8 M GdmCl. The dotted magenta line is average deuteron uptake of W42R in 1.8 M GdmCl. B) Subtractive deuteration plot of W42R and V75D.



Supplemental Figure 2.3: A) Uptake plots comparing V132A γDC$_{CTD}$ and wild-type γDC$_{CTD}$. The dotted magenta line is average deuteron uptake of V132A in 4.8 M GdmCl. B) Subtractive deuteration plot of V132A γDC$_{CTD}$ and wild-type γDC$_{CTD}$. V132A γDC$_{CTD}$ has minor loss of protection in peptides near the site of mutation, but otherwise closely resembles the wild-type γDC$_{CTD}$.

Chapter 3. Probing the mechanical unfolding of a ubiquitinated protein using optical tweezers

3.1 Introduction

This chapter focuses on the effect of ubiquitin on the force-induced unfolding of a model protein, barstar. After an introduction to the question and an orientation to the experimental approach used, I first describe my attempts to generate samples for this analysis, detailing both successes and failures. I next report on my general characterization of ubiquitinated barstar using force-ramp experiments. I settled on using force-jump unfolding to characterize unfolding behavior of one single-lysine variant. I found that ubiquitination of barstar at position 60 both increases the force-induced unfolding rate and potentially changes the pathway by which barstar unfolds, underscoring the value in studying protein systems using physiologically relevant perturbations.

3.1.1 Ubiquitin is a post-translational modification necessary for proteasomal degradation

Ubiquitin is a common post-translational modification. It plays crucial signaling roles in various regulatory nonproteolytic cellular activities, including vesicle trafficking, protein binding partner recruitment, endocytosis, and enzymatic activity (80–82). However, the most widely recognized and most well-understood function of ubiquitin modification is to mark substrate proteins for degradation by the 26S proteasome (83, 84) (Figure 3.1).

Targeted degradation of eukaryotic proteins is primarily undertaken by the ubiquitin-proteasome system (UPS) (85, 86). Proteins are marked for proteolytic degradation by the 26S proteasome via the attachment of ubiquitin by the E1/E2/E3 ubiquitination machinery (87). The proteasome's 19S regulatory particle is responsible for recognizing ubiquitinated substrates, deubiquitinating them, mechanically unfolding them in an ATP-dependent fashion, and subsequently translocating the now unfolded polypeptide into the 20S core particle. Proteolysis is performed within a central cavity of the core particle that contains proteolytic sites.

The proteasome must process a wide variety of protein substrates with highly diverse chemical and structural parameters without sacrificing selectivity. Considering the vital function of the UPS in governing protein activity, it comes as no surprise that malfunctions in UPS components are associated with the development of various human pathological conditions. These conditions encompass a range of maladies, including cardiovascular diseases, multiple cancers, and neurodegenerative disorders (88). Specificity of proteasomal engagement is therefore crucial.

Notably, the regulatory particle requires engagement of an unstructured region of the substrate in order to initiate degradation (89, 90). While this region may occur on the N- or C-terminus or even in an unstructured loop, 30% or more clients of the proteasome lack such an unstructured region (91).

Experimental characterization of ubiquitin-mediated changes in the energy landscape of barstar has demonstrated that ubiquitin has site-specific impacts on protein energetics, demonstrating the ability to induce unstructured regions and promote engagement by the proteasome (92). These effects of ubiquitin on the energy landscape of substrate proteins are site-specific in both degree and mechanism of impact, and can range from negligible to substantial (93).

There remain many important open questions relating to the diverse effects of ubiquitination on its substrates. How are degradative and regulatory effects distinguished? For proteins lacking disordered regions, how does ubiquitination promote targeting and engagement by the proteasome? How do subtle effects on native conformations lead to significant changes in degradation outcomes?

Finally, what is the significance of the mechanical context of the proteasome? Many thermodynamic and kinetic studies have used global perturbants such as chemical denaturants, temperature, and pH to probe protein energy landscapes in great detail. However, it is not clear that parameters obtained by these experiments are germane to the dynamics and high-energy states sampled when a protein experiences mechanical force along a given vector. This is a critical distinction, given the number of physiological contexts in which some degree of mechanical force is relevant.

Destabilizing ubiquitin modifications can increase the rate of substrate degradation by the proteasome *in vitro* (92). As unfolding and degradation by the proteasome is a mechanical process, this suggests that ubiquitination may alter substrate stability in a way that is particularly germane to force-unfolding, thus assisting in the preparation of substrates for proteolysis.

Given the highly context-dependent effects of ubiquitination on the energy landscape of a modified protein, I asked how ubiquitination can impact a protein's mechanical unfolding behavior.

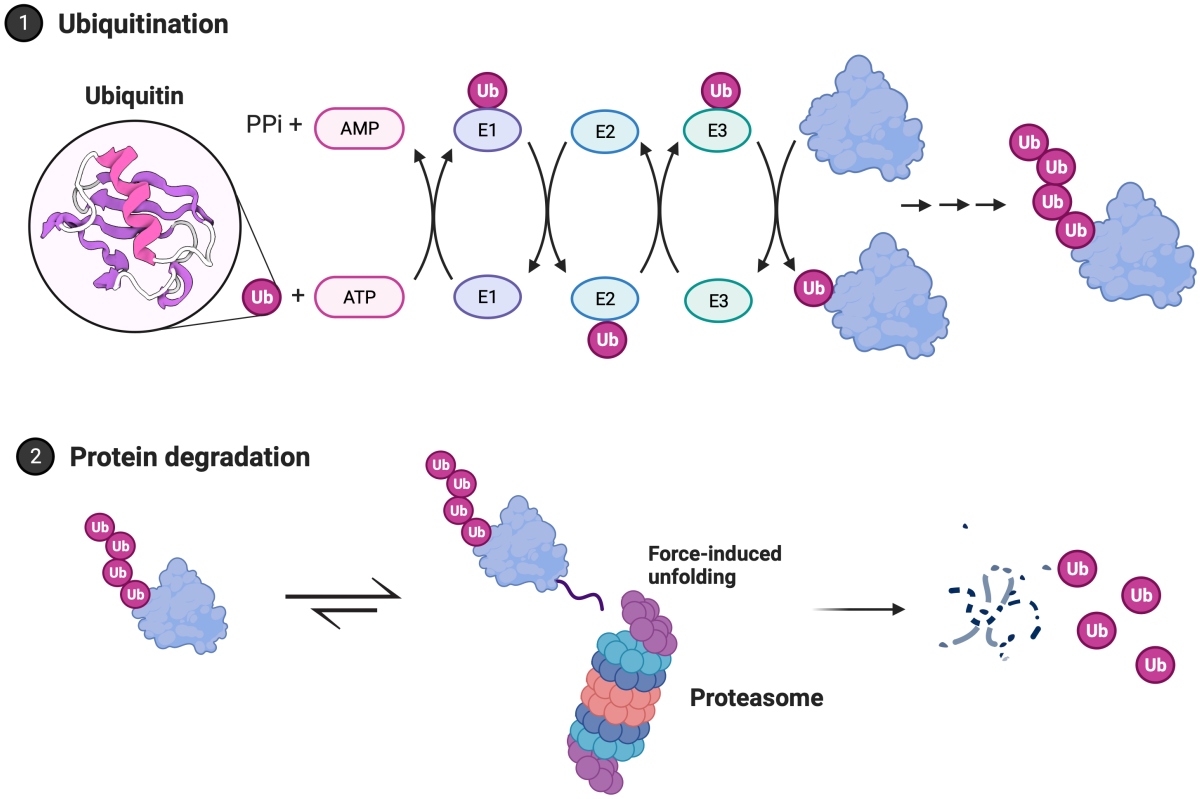


Figure 3.1: Substrates destined for degradation by the proteasome are ubiquitinated via a complex enzymatic cascade. Subsequently, in the canonical model for proteasomal degradation of ubiquitinated substrates, the 26S proteasome engages an unstructured region of the target. The substrate is then mechanically unfolded and translocated into the 20S core particle for proteolytic degradation. Figure adapted from “Ubiquitin Proteasome System”, by BioRender.com (2023).

3.1.2 Force is a biologically relevant perturbant

Force plays a fundamental role in biological processes (94). Mechanical perturbation commonly occurs in the function of biological systems: cell adhesion proteins, muscle proteins, extracellular matrix proteins, and more undergo force routinely. Additionally, virtually every cellular process is dependent on properly regulated protein degradation; the proteasome unfolds proteins via a forceful mechanochemical translocation mechanism (95).

Force deforms the protein energy landscape in ways distinct from chemical and thermal denaturation (96), as the shifts in both the high-energy states sampled and the associated energy barriers depend on the method of perturbation used. Traditional

protein folding experiments rely on global changes in solution to evince unfolding: pH, temperature, or chemical denaturants such as urea or guanidium chloride. The details of how these perturbants alter the energy landscape depending on the specifics of the target protein have been well-characterized (51). However, measurements of a protein's mechanical elasticity and the anisotropic stability barriers experienced under force are usually not as well understood.

Given the mechanical physiological context of many proteins, this is a critical factor. The parameters derived, for example, from urea melts may be accurate in the context they are attained yet may not translate in terms of physiological relevance. As one example of how force is a conceptually distinct means of denaturation, external force drastically reduces the entropy of the unfolded state by stretching and extending a polypeptide chain. Therefore, chain entropy plays a less important role in mechanical unfolding and a single reaction coordinate, such as end-to-end distance or number of native contacts, should adequately describe the unfolding process (62).

Force is known to influence protein unfolding pathways. In the case of srcSH3, a small, single-domain model protein, presence or absence of force may bias unfolding flux through one of three distinct pathways, depending on the degree and orientation of force applied (97). Force can therefore not only modulate barrier heights, but also change the identity of the intermediate states populated.

3.1.3 Single-molecule force spectroscopy can be used to exert and measure piconewton-scale forces on molecules

Single-molecule force spectroscopy provides a direct measurement of a protein's response to mechanical force along a specified reaction coordinate with a physical and intuitive interpretation: extension length. As a single-molecule method, it also has the advantage of probing stochastic behavior, alternate folding trajectories, and rare events impossible or very difficult to characterize in ensemble experiments. It has high spatial resolution, high detection sensitivity for rare states, high time resolution, and can also be used to measure thermodynamic and kinetic parameters in multiple redundant ways (98).

Force-dependent conformational changes can be monitored by an optical trap, a single-molecule force spectroscopy technique that is ideally suited for lower force measurements. Optical traps or optical tweezers rely on the use of a highly focused laser to apply force to manipulate small particles. Ashkin demonstrated that tightly focused laser light can be used to stably hold microscopic particles (99). The particle suspended in a trap acts as a refractive object. Conservation of linear momentum dictates that when the incident light changes momentum upon refraction by the particle, the particle must also change its momentum. Because forces exerted by, or on, the bead result in deflection of the trapping beam, which can be directly measured by a

position-sensitive photodetector, both the force and displacement that molecule experiences may be measured.

Molecules of interest (protein, DNA, etc.) are typically too small to interact with the trapping light and hence must be tethered to a micron-sized bead via a linker, or “handle" molecule (100). Generally, dsDNA handles are used, as their stability and behavior under stretching forces have been well-characterized (101, 102). In this experimental setup, two DNA handles are used to tether individual molecules of a protein of interest between two micron-sized polystyrene beads; one bead is held in place by a pipette, while the other is manipulated by the optical trap, permitting the protein to be stretched and relaxed. The low spring constant of an optical trap (0.005-1 pN/nm) allows measurement at low forces (1-100 pN) and low loading rates (1-100 pN/s), with sensitivity on the order of ~0.1 pN and spatial resolution on the order of a few nanometers (103). This permits the probing of kinetic regimes close to equilibrium conditions and physiological forces, enabling direct observation of fluctuations between different conformations.

The precise range of mechanical force generated by the proteasome while translocating substrate has not been measured. However, the stall and translocation forces of ClpX, a bacterial protein unfoldase, have been measured on the order of ~20 pN (104). If this is in the same realm of force experienced by ubiquitinated proteasome substrates, then optical tweezers are a perfect method to probe mechanical stabilities in a physiologically relevant context.

3.1.4 Ubiquitin attachment proceeds via an enzymatic cascade

Ubiquitin is a 76-amino acid, 8.5 kDa, highly conserved eukaryotic protein that is typically attached to a substrate protein via an isopeptide bond between ubiquitin’s C-terminus and the ε-primary amine of a lysine residue on the target protein. From there, ubiquitin may be subjected to further modification, including polyubiquitination, which can result in complex signals referred to as the “ubiquitin code,” for a number of distinct cellular outcomes (80, 105).

This attachment proceeds via a highly regulated enzymatic cascade involving three types of enzymes: E1, E2, and E3 (106) (Figure 3.1). First, an E1 activating enzyme employs ATP to activate the C-terminus of ubiquitin via the formation of a high-energy thioester bond. Next, the activated ubiquitin is transferred to a cysteine residue of an E2 ubiquitin conjugating enzyme via a transthiolation reaction. Finally, an E3 ubiquitin ligase recruits E2 enzymes and facilitates the transfer of ubiquitin to a protein substrate, acting either as a covalent intermediate or as a scaffold (105). Subsequently, ubiquitin itself may be ubiquitinated on its seven lysine residues or on its N-terminus, yielding a diverse array of ubiquitin chain lengths and topologies on a target protein. E3 is the primary controller of substrate identity, while both E2 and E3 play roles in determining chain linkage type and length (106).

High-purity mono-ubiquitinated substrates may be produced *in vitro* via a biochemically reconstituted version of this enzymatic cascade (92). By use of a maltose binding protein (MBP) scaffold coupled to a PPPY recognition motif for E3-substrate engagement, ubiquitin may be appended to lysine residues within structured domains with native isopeptide linkages.

In order to be sure of substrate homogeneity, substrates containing only one lysine and octadimethylated ubiquitin (dimethylation at its N-terminus and at all its native lysine residues) can be employed to ensure that the resulting ubiquitinated product a) is uniquely ubiquitinated at one residue and b) is mono-ubiquitinated rather than heterogeneously poly-ubiquitinated.

3.1.5 Scope and summary

In the work presented in this chapter, I sought to address the significance of the mechanical context of the proteasome by investigating how ubiquitination affects the mechanical unfolding pathway of a protein.

I aimed to tackle the following questions. Does ubiquitination affect the mechanical stability of a protein? Does it affect the rate of force-induced denaturation? Can ubiquitination change the unfolding pathway by altering unfolding intermediates occupied or influencing the distance to the transition state?

First, I needed to produce mono-ubiquitinated protein substrates of very high purity. Next, I needed to develop and employ a handle attachment strategy that worked successfully with both ubiquitinated and non-ubiquitinated substrates. Finally, I needed to employ force-ramp and force-jump experiments on both ubiquitinated and non-ubiquitinated substrates to characterize and compare the force-unfolding landscape. The main challenges to overcome included systemization of handle attachment and successful attachment of handles to ubiquitinated substrates, which are challenging to work with.

As a target protein, I employed a well-characterized model protein, barstar from *Bacillus amyloliquefaciens*, in which all lysines but one were mutated to arginine in order to generate different single-lysine variants (107). Not only has barstar been well-characterized by bulk biophysical methods (108, 109), but ubiquitination has also been shown to have site-specific consequences on its stability and susceptibility to proteasomal degradation (92).

As detailed below, I demonstrate that ubiquitination of barstar at K60 not only changes its force-induced unfolding rate, but also influences the responsiveness of unfolding rate to force, implicating a change in unfolding pathways.

3.2 Handle attachment strategies

In order to study a protein in the optical trap, it must be covalently conjugated to a dsDNA "handle". Originally, I set out to do this using the previously described approach of engineering cysteines at the specific sites for handle attachment and using DTDP (2,2′-dithiodipyridine)-mediated thiol chemistry (100). This chemistry has been employed successfully in the past, but I encountered several issues when attempting to apply it here. One issue is the contamination of successfully generated DNA handle-protein-handle chimeras with the competing handle-handle product. This problem may be allayed somewhat by the use of a histidine tag and a Ni^{2+}-NTA pulldown step to enrich the desired reaction product, but this off-pathway product still substantially reduces the efficiency of the reaction. There is a similar issue with the possibility of contaminating poly-protein constructs. Finally, I encountered issues in the degree of success of the reaction when long (500-1000 bp) handles are employed in the direct reaction mechanism, which appears to reduce success rates of functionalization.

All in all, handle attachment is troublesome, especially because gel-based assays are frequently unreliable for unambiguous positive identification of DNA-protein chimeras. I identified several troubleshooting hurdles with respect to collecting single-molecule data. While I cannot prescribe a strategy to fix these problems unambiguously, I provide ideas and tips to help overcome them in Table 3.1, and I also provide details on both successful and unsuccessful handle attachment strategies.

Tether scarcity: few tethers catching	Sample freshness	Always store at -80˚C; do not store for longer than ~3 months; perform final ligations fresh every day
	Not enough sample added	Increase amount of protein-handle chimera incubated with anti-dig beads or increase time of incubation
	Final oligonucleotide-handle ligation unsuccessful	Check reagent freshness
	Oligonucleotide-protein crosslinking unsuccessful	Change ratio of handles:protein used during attachment reaction
		Make sure that protein is fully reduced prior to attachment reaction, and move quickly once

		reducing agent is removed
Rip scarcity: very few or no protein unfolding transitions observed on otherwise successful tethers	Too much handle homo-dimer formation leading to scarce protein-DNA formation	Change handle strategy to preclude the possibility of handle-handle chemistry (ie, use bifunctional groups such as SMCC or GMBS)
	Protein that is already unfolded or otherwise of low quality	Verify folded protein by an orthogonal technique
	Not enough enrichment of sample	Use His tag/Ni^{2+}NTA or other pulldown technique to enrich final product
	Poor reduction of protein prior to functionalization	Increase time or temperature of reduction
Tether fragility: tethers breaking at low forces	Poor interaction with beads	Vary amount of protein-handle chimera incubated with anti-dig beads or vary time of incubation
Artifacts: poly-protein constructs, messy tethers, wrong-sized rips	Difficulty identifying unfolding transitions	Careful comparison with samples of known quality
	Contamination in the chamber	Replacing the chamber; washing anti-digoxigenin beads with buffer
	Beads too sticky	Passivating streptavidin beads by vortexing with 2 mg/mL BSA; washing anti-digoxigenin beads with buffer
	Multiple tethers caught	Reduce amount of protein-handle chimera incubated with anti-dig beads or reduce time of incubation

Table 3.1: Common technical problems and potential troubleshooting solutions in tweezers sample preparation and handling.

3.2.1 Negative result: sortase

To avoid thiol chemistry altogether, I attempted an alternative scheme for handle attachment utilizing sortase-able handles (Figure 3.2). Towards that goal, I used a different model system – I27 from titin (110). Sfp phosphopantetheinyl transferase was used to site-specifically label an N-terminal short peptide tag with a DNA oligo that had been covalently conjugated to CoA via a terminal thiol (111). The oligo, which contains a short overhang, was then ligated to a 1 kbp DNA handle modified with 5'-digoxigenin. Sortase-mediated transpeptidation is a method for site-specific addition of a polyglycine peptide to a sortase tag recognition sequence (LPXTG) at the terminus of a protein (112). A 1 kbp sulfhydryl-modified DNA handle modified with 5'-biotin was labeled with a short peptide (GGGK) via sulfo-GMBS, a sulfhydryl-to-amine crosslinker. Subsequently, sortase was used to functionalize the C-terminus of I27, which was modified with a sortase-recognition motif, to this handle.

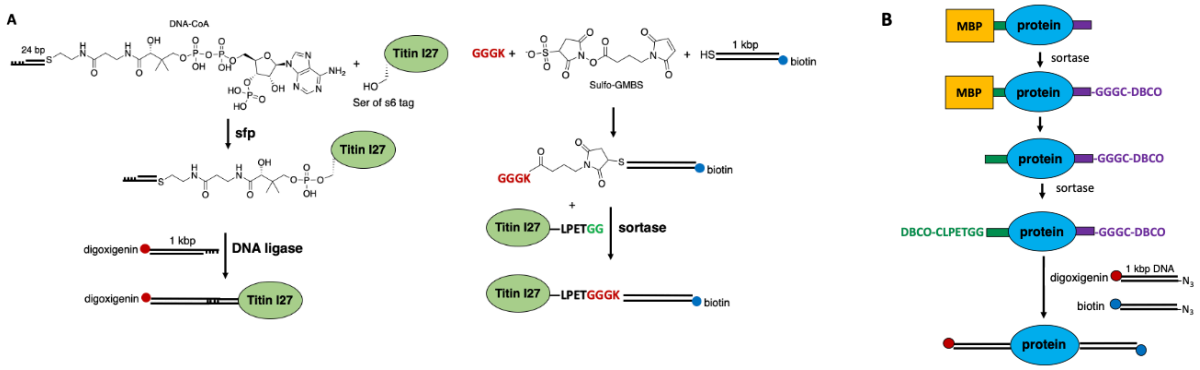


Figure 3.2: Two unsuccessful strategies for covalently attaching DNA handles to both termini of a substrate protein (in this case, titin I27) utilizing short peptides and sortase chemistries. A) N-terminal attachment utilizing sfp chemistry and C-terminal attachment utilizing sortase chemistry. B) DBCO-maleimide modification of short peptides followed by N- and C-terminal sortase attachment, and subsequent click chemistry attachment to azide-modified handles.

Despite multiple attempts at using gel-based assays, I was never able to adequately show functionalization of peptide-DNA handles to substrate proteins. It is possible that the length of the DNA handle prevented adequate interaction with the sortase active site. Therefore, I attempted a modification of this protocol that relied on 1) modifying short peptides with DBCO-maleimide, 2) using N- and C-terminal sortases to attach these peptides to N and C-termini of a substrate protein, and 3) using click chemistry to attach the substrate protein to azide-modified DNA handles (113) (Figure 3.2B). Each successive reaction step reduces the overall probability of the entire reaction taking place successfully and increases the risk of the substrate protein unfolding or misfolding. It was difficult to identify which step was failing and therefore I aborted this approach.

3.2.2 Positive result: GMBS with short handles

I next attempted handle attachment to barstar substrates by directly using GMBS (N-γ-maleimidobutyryl-oxysuccinimide ester)-mediated thiol chemistry. While this is conceptually similar to DTDP chemistries in that it relies on chemical functionalization of reduced and solvent-exposed cysteines on the target protein, GMBS has a reactive NHS ester group in addition to its maleimide group. This bifunctionality means that GMBS can be used as a sulfhydryl-to-amine crosslinker and thereby exclude the possibility of sulfhydryl-to-sulfhydryl (handle-handle or protein-protein) chemistries.

Direct reaction with full-length handles proved inefficient, so I made use of a protocol that crosslinks short NH2-modified DNA oligonucleotides (20 bp) to protein substrates prior to ligation with longer handles (personal communication, Robert Sosa (Figure 3.3). This was more successful. I found that inclusion of an N-terminal His6 tag for purification from unattached oligos is crucial to the success of this method. Enrichment by reliance on amylose binding to MBP following ubiquitination was unsuccessful. Extremely thorough reduction of protein substrates followed by careful desalting so as not to reduce yield was also found to be critical. Altering ratios of DNA to protein during the final attachment step was found to be helpful in increasing overall yield, though this effect seemed sample dependent.

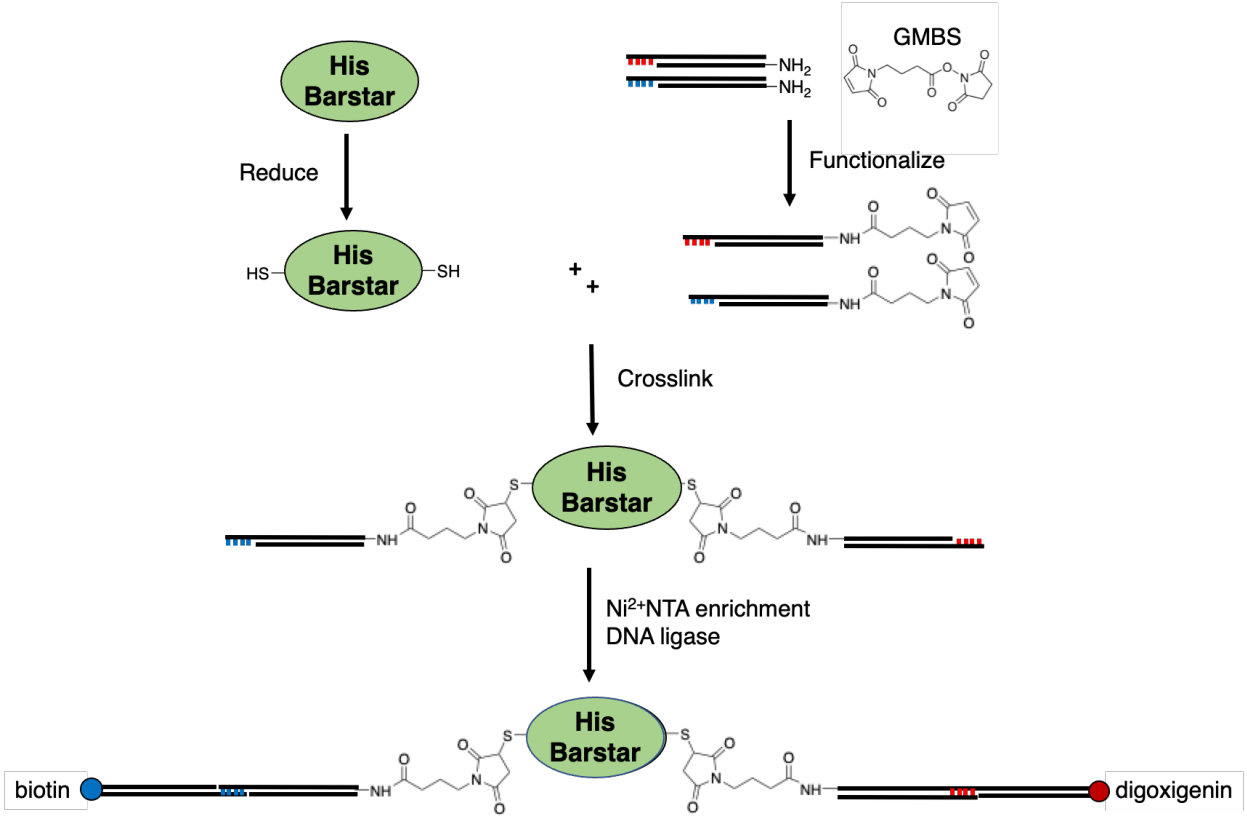


Figure 3.3: Successful strategy for covalently attaching DNA handles to two cysteines of a substrate protein utilizing GMBS chemistry. GMBS is used to functionalize two short amine-modified DNA oligos with short overhangs, which are subsequently

crosslinked to a target protein containing two reduced cysteine residues. The DNA-handle chimera can then be ligated to longer handles with matching overhangs.

3.3 Results

Barstar with a single lysine at position 60 and all other lysines mutated to arginine, termed barstarK60, was engineered with two cysteines, one appended to the N-terminus (position 0) and one native cysteine at position 82. A second native cysteine at position 23 was mutated to alanine. This substrate was mono-ubiquitinated. 558-bp DNA handles were attached at each cysteine in both the mono-ubiquitinated substrate (termed Ub barstarK60) and the non-ubiquitinated substrate (barstarK60). Force-ramp and force-jump experiments were performed on each construct to assess mechanical unfolding properties (Figure 3.4).

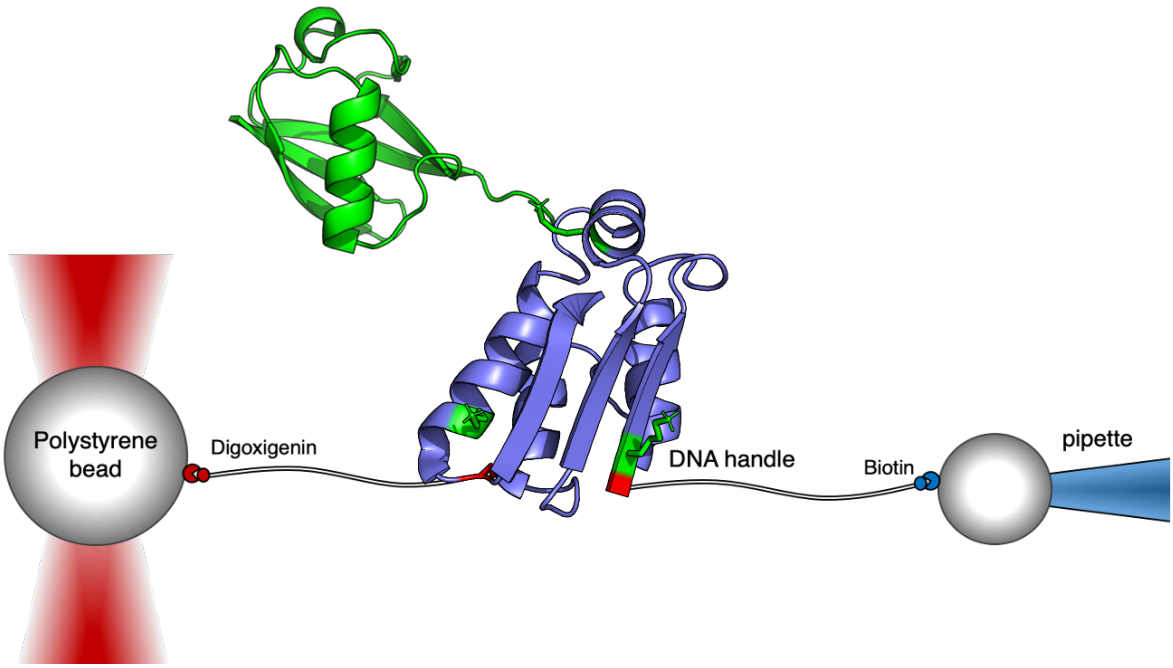


Figure 3.4: Schematic for optical trapping experiment (not to scale). In this case, mono-ubiquitinated (green structure, PDB: 1UBQ (114)) barstarK60 (blue structure, PDB: 1bta (115)) is bound between two micron-sized beads by two DNA handles (size 558 bp) which are covalently attached to two different sites on the protein via thiol-GMBS chemistry at two cysteine residues (C0 and C82, red). One handle tethers the sample to a bead held in the optical trap via a digoxigenin-antibody interaction; the other handle tethers to a bead held by suction to a pipette via a biotin-streptavidin interaction. Green sites indicate two other single-lysine variants prepared, K2 and K78.

3.3.1 Force-extension curve analysis

In a force-ramp experiment, the two beads are moved apart at a constant velocity, which in turn continuously increases the force exerted on the molecule between them. Prior to protein unfolding, the handles are stretched, generating an elastic increase in extension according to the wormlike chain model (116). Once the protein is no longer mechanically stable, it unfolds in a sharp transition that appears as a sharp increase in contour length accompanied by a drop in force, termed a “rip”. Following unfolding, additional extension continues with increased force, with polymeric behavior now determined by both the unfolded portion of the protein and the DNA handles. At high forces (F = 60-70 pN), the force plateaus and extension rises quickly due to DNA overstretching (117). Upon lowering the force, a refolding event may be observed, a transition characterized by a reduction in contour length and a jump in force. Both of these events are stochastic.

This is a non-equilibrium experiment, often with significant hysteresis observed between unfolding and unfolding trajectories; however, collection of distributions of folding and refolding forces can yield equilibrium information about the folding-unfolding process by employing fluctuation theorems to extract free energy from the overlaps between two work distributions (118, 119).

Force ramps can also be used to obtain other fundamental information about the unfolding process. The observation of multiple transitions corresponds to multiple unfolding intermediates. Each state in a force-extension curve should yield part of a curve following an appropriate polymer elasticity model, which can reveal states with different unfolded contour lengths corresponding to different numbers of amino acids unfolded. Moreover, force ramps cover a broad range of forces and are thus useful to detect transitions in proteins where the mechanical stability is unknown.

I employed force ramps for an initial characterization of barstarK60 under force, pulling at a rate of 100 nm/s from 3 to 65 pN generate force-extension curves. No more than twenty rips per unique molecule were analyzed, and at least five (usually much more) unique molecules were analyzed per protein (Figure 3.5).

I observe at least one, and possibly multiple, unfolding intermediates, distinguished as multiple rips observed on the same force extension curve. These intermediates are not necessarily obligatory, as they are not always observed. This is in concordance with previous work in bulk experiments demonstrating that barstar can populate multiple kinetic unfolding intermediates and has more than one unfolding pathway (108, 109). I chose not to carry out a detailed analysis of these force-ramp experiments as the nonlinearity of the trap stiffness in our mini-tweezers in the higher force regimes precludes direct fitting of these curves to a wormlike chain without substantial correction.

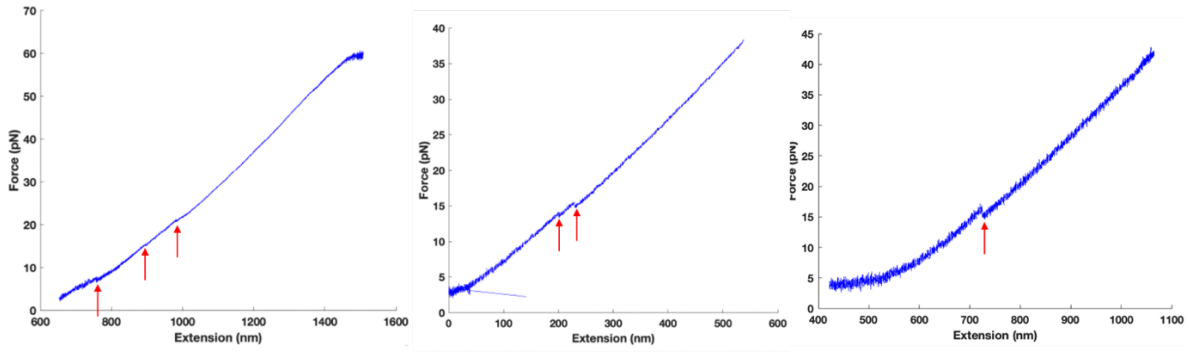


Figure 3.5: Force ramps of barstarK60. BarstarK60 can unfold via one or more unfolding intermediates, as indicated by the observation of between one and three rips (indicated by red arrows).

In spite of this, I was able to analyze the molecular length involved in a given unfolding event, Δx, which can be obtained from the difference between the two bead positions. I fit the force and extension of each separate rip to a WLC model to calculate the consensus contour length and compare it to a predicted contour length based on the number of amino acids which should be involved in a complete unfolding event (Figure 3.6). All force ramp transitions for barstarK60 in which only one unfolding event was observed was fit to a WLC according to the following equation (120):

$$\frac{x}{L_0} = \frac{4}{3} - \frac{4}{3\sqrt{\frac{FP}{k_BT}+1}} - \frac{10e^{\sqrt[4]{900\frac{k_BT}{FP}}}}{\sqrt{\frac{FP}{k_BT}}\left(e^{\sqrt[4]{900\frac{k_BT}{FP}}}-1\right)^2} + \frac{\left(\frac{FP}{k_BT}\right)^{1.62}}{3.55+3.8\left(\frac{FP}{k_BT}\right)^{2.2}} \qquad (1)$$

This yielded a contour length of 27.5 ± 0.4 nm for barstarK60, with 129 rips analyzed. This is in relatively good agreement of the predicted contour length of 28.7 nm, though it is slightly too low. This underestimate is likely due to the known presence of unfolding intermediates, as even though effort was taken to only include singular rips in this analysis, some of the rips included in this fit might not represent full unfolding events. Ubiquitinated barstarK60 yielded similar results, with no significant differences in contour length or the incidence of intermediate behavior (L = 31.4 ± 1.2 nm, n = 17 rips). The lower sampling size for ubiquitinated barstarK60 is due to difficulty in collecting higher force data without losing the tether; due to the challenges in data collection, I chose to focus on force-jump rather than on force-ramp experiments for ubiquitinated samples.

Unfolding force histograms were also generated for both ubiquitinated and non-ubiquitinated samples, demonstrating significant overlap between the two populations (Figure 3.6). The low number of samples collected for ubiquitinated barstarK60 precludes me from drawing any conclusions about differences in unfolding force histograms. Because refolding transitions took place at forces that were often too low to detect unambiguously, I was unable to collect unfolding force distributions.

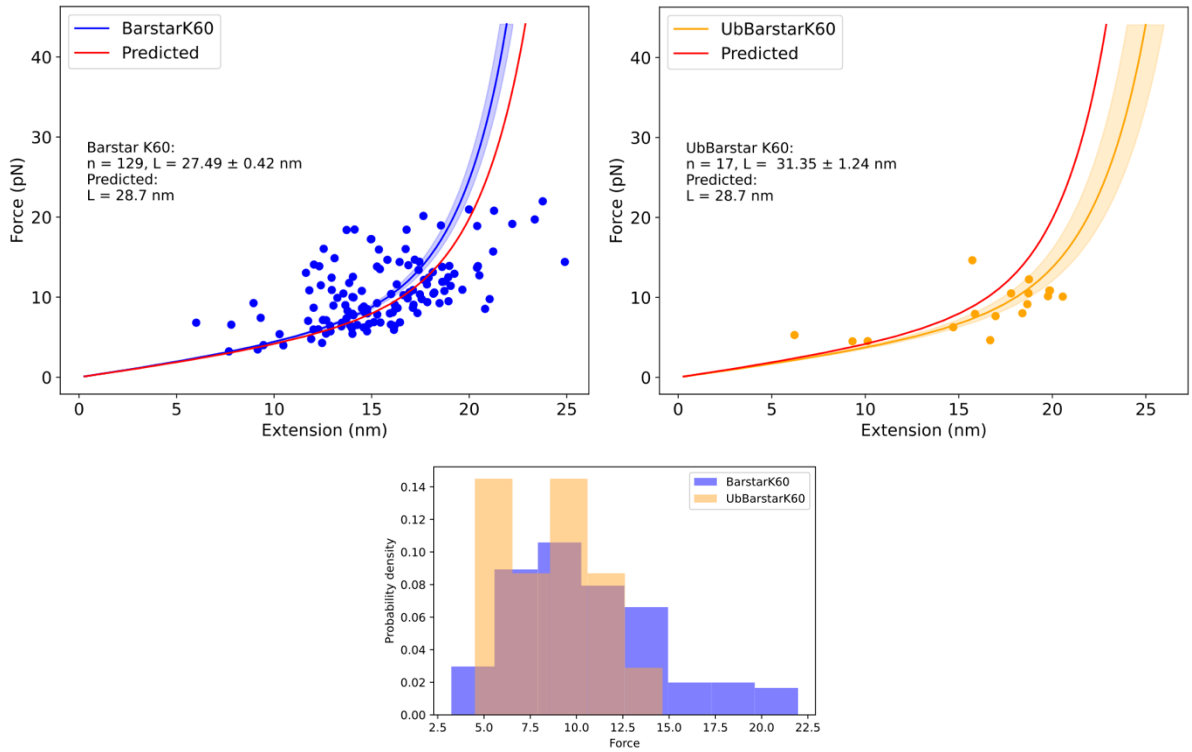


Figure 3.6: Force rip distributions and unfolding force histograms for barstarK60 (blue) and Ub barstarK60 (orange). Relatively close agreement to predicted contour length indicates that unfolding transitions correspond to full unfolding of barstarK60.

3.3.2 Force-jump analysis

A force-jump experiment, in contrast to a force-ramp experiment, relies on an abrupt transition between two constant forces. In a force-jump experiment, a molecule is equilibrated at a low, constant force, and then is “jumped” to a higher constant force (121). The dwell time of the molecule in its folded state before it undergoes an unfolding transition can then be measured. By repeating this experiment and measuring a distribution of dwell times, the average lifetime of the folded state can be determined. In this manner, a force jump experiment permits direct measurement of folded lifetimes as a function of the magnitude of the maintained constant force.

The simplest model to describe how an applied constant force will impact the unfolding rate constant is one in which free energy is linear with respect to force, as in Bell's model (122):

$$k_u(F) = k_m k_u^{0pN} e^{\frac{F\Delta x^{\ddagger}}{\mathrm{k_B T}}} \quad (2)$$

where k_u is the observed unfolding rate constant, k_m includes the contributions of the experimental system to the observed rates, including beads, handles, and trap stiffness, k_u^{0pN} is the intrinsic rate constant in the absence of force, F is the applied force, $\Delta x^{\ddagger}$ is the distance along the reaction coordinate to the transition state, k_B is the Boltzmann constant, and T is temperature.

Thus, the unfolding force and the natural log of the rate of unfolding have a linear relationship that can be described as follows:

$$\ln\left(k_u(F)\right) = ln(k_m) + \ln\left(k_u^{0pN}\right) + \frac{F\Delta x^{\ddagger}}{k_B \mathrm{T}} \quad (3)$$

In this case, the end-to-end extension of a molecule serves as the order parameter. As long as states are well-separated along this reaction coordinate, this can serve as an informative parameter to distinguish different states (102). The slope of the natural log versus force is independent of the spring constant of the trap and system and can be related to $\Delta x^{\ddagger}_{\text{unfolding}}$, the distance along the reaction coordinate from the folded state to the force-induced transition state. Determination of the position of the transition state along the reaction coordinate can be helpful in distinguishing unfolding pathways from one another (123). Moreover, deviations from linearity can suggest unfolding via different transition states in different force regimes (97).

Using force feedback, barstarK60 was jumped and held at either folding (3 pN) or unfolding (7-20 pN) forces until a transition in length was observed (Figure 3.7). Each individual molecule was switched between folding and unfolding regimes no more than twenty times per force studied, and at least five (usually much more) unique molecules were assessed per protein to confirm consistency of kinetic behavior between tethers. Unfolding jumps were then pooled for distribution analysis.

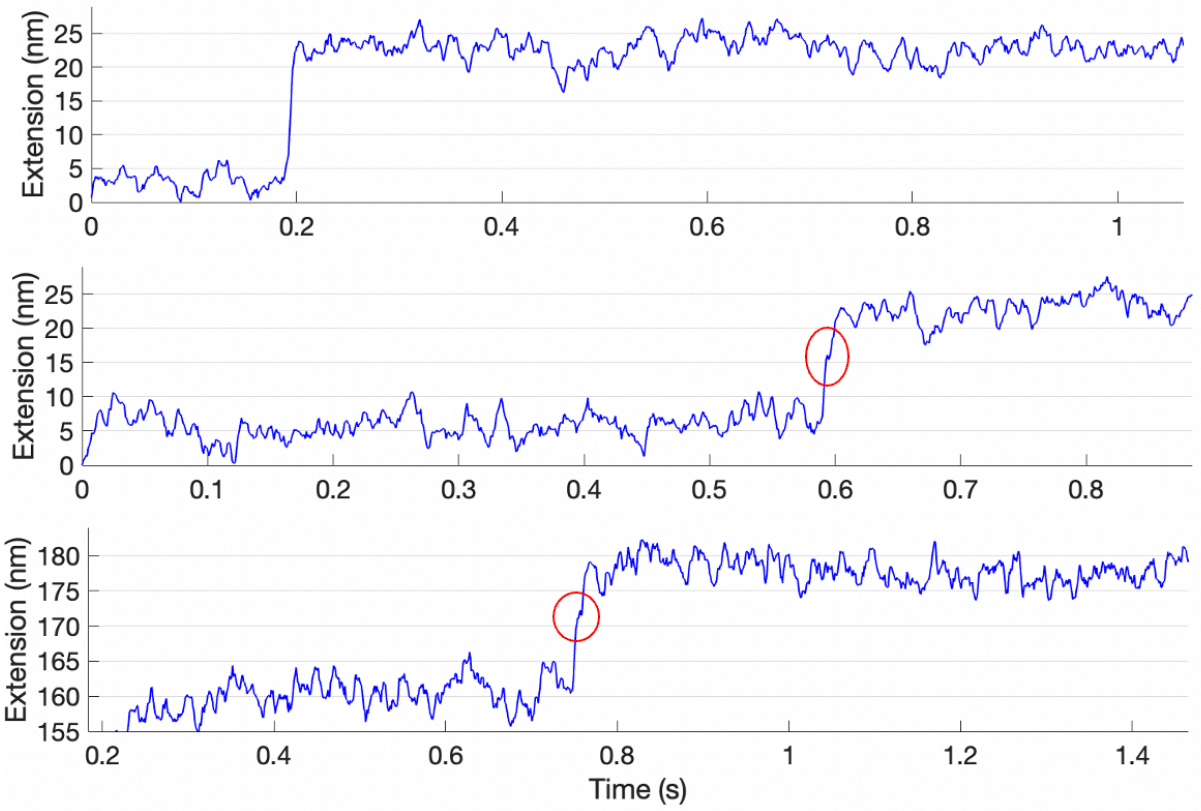


Figure 3.7: Example unfolding traces of barstarK60 in force jump experiments. At time = 0, force was jumped from low (3 pN) to high (15 pN in this case). Short-lived unfolding intermediates were sometimes observed, circled in red.

Unfolding rates were calculated at each force by fitting to a kinetic model wherein the probability of each molecule unfolding at a given time was presumed to obey an exponential distribution,

$$p(t) = k_u e^{-k_u t} \tag{4}$$

where p(t) is the probability of observing a transition at a given time, k_u is the unfolding rate, and t is time. To verify that distributions were consistent with the underlying populations, I chose half the samples at random and repeated the fits, finding consistent results. Distributions were only fit if the number of observed rips exceeded 15 (usually the case); otherwise, the mean rate was used (Figure 3.8).

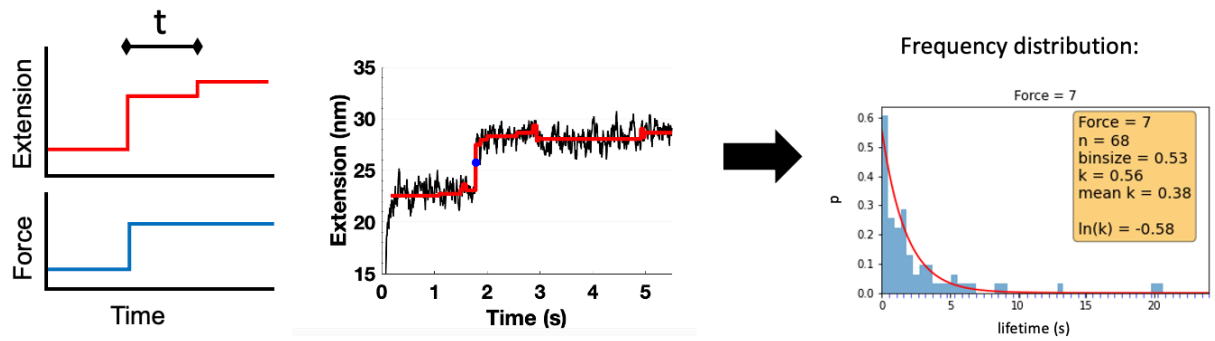


Figure 3.8: Schematic of force jump experiments. Samples are held at low force and then jumped to high force. Distributions of unfolding lifetimes at each high force can be fit to exponential rates.

I observe that, in accordance with force ramp observations, barstar K60 has the ability to unfold via an unfolding intermediate consisting of approximately half of the protein's full contour length. This intermediate is either easier to observe (due to slower unfolding rates) or more frequently populated at lower forces. Ubiquitination does not seem to significantly influence the incidence of observing the intermediate (Figure 3.9).

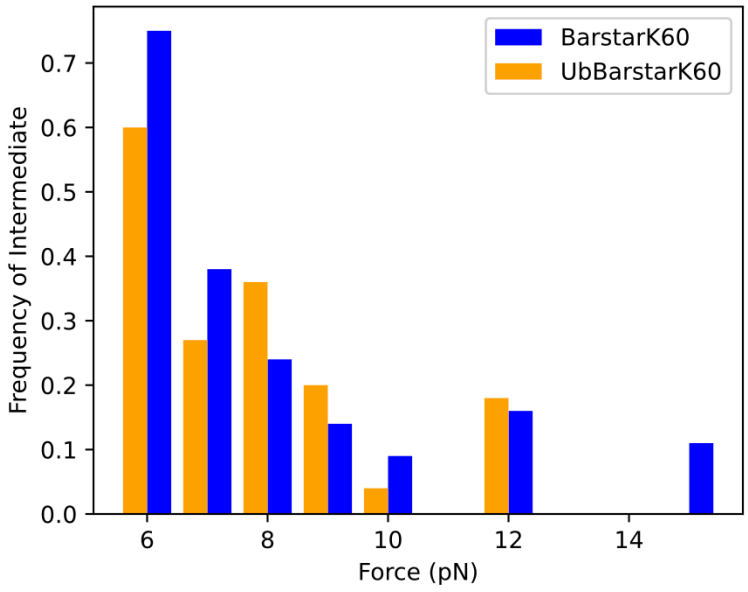


Figure 3.9: Frequency of observing an unfolding intermediate in force jump experiments for barstarK60 (blue) or Ub barstarK60 (orange). Ubiquitination does not seem to dramatically influence the probability of observing an intermediate.

I find that the natural log of the unfolding rate of barstarK60 is indeed linear with respect to unfolding force, with $\Delta x^{\ddagger}_{unfolding}$ = 1.0 ± 0.11 nm (r^2 = 0.965) (Figure 3.10). The natural log of the extrapolated rate constant of unfolding at zero force is -2.53, which is not dissimilar to the extrapolated rate found in bulk studies by Zaidi *et al.* (108). The value I report is not expected to be directly comparable to unfolding rates measured by chemical denaturation, both because the presence of beads, handles, and the trapping apparatus in force spectroscopy experiments are known to alter the unfolding rate (the $\ln(k_m)$ term), and because linear extrapolation from high denaturant may not match unfolding behavior at low denaturant (especially when unfolding intermediates are known to be involved) (124).

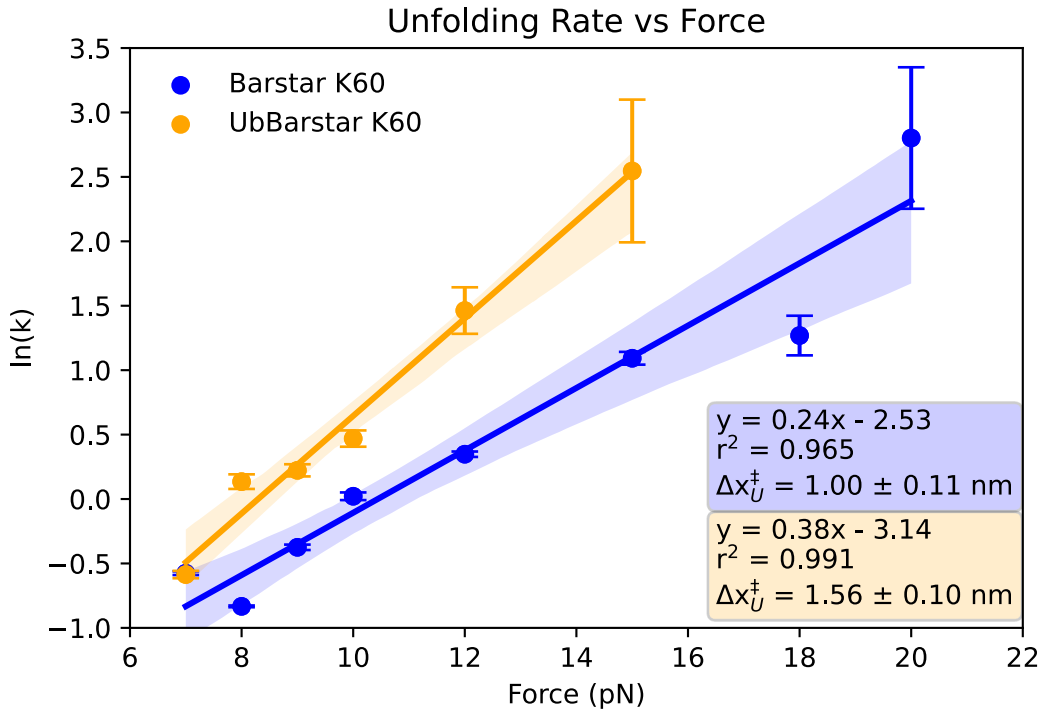


Figure 3.10: Force dependence of mechanical unfolding rates of barstarK60 (blue) and Ub barstarK60 (orange). Linear fits are to Bell's model, with 95% confidence intervals indicated by shaded regions. Slopes of each line are statistically distinct by a one-way F-test ($p = 3.7 \times 10^{-5}$). Error bars represent the standard error of each exponential fit.

Upon ubiquitination, I find that Ub barstarK60 unfolds faster at higher forces, indicating that ubiquitination accelerates mechanical unfolding (Figure 3.10). Fascinatingly, I also find that ubiquitination changes the slope of the Bell relation, indicating that ubiquitination makes the rate of unfolding more sensitive to force. I verified that the data are much better fit by assuming a different linear relationship between ln(*k*) and force for Ub barstarK60 than for barstarK60 ($p = 3.7 \times 10^{-5}$ by one-way F-test). Therefore, the linear relationship between ln(*k*) and F for Ub barstarK60 has a steeper slope than its

non-ubiquitinated counterpart, and the distance to the transition state is increased from $\Delta x^{\ddagger}_{unfolding}$ = 1.0 ± 0.11 nm to 1.56 ± 0.10 nm (r^2 = 0.991). The extrapolated ln($k_{unfolding}$) at 0 pN is -3.5, implying a crossover at low force. It is possible that there is nonlinear curvature at low force. This implies that barstar K60 not only unfolds faster when ubiquitinated, but it may also be unfolding via a different transition state and therefore an alternative unfolding pathway.

3.4 Conclusions, limitations, and future directions

The idea that ubiquitination is sufficient to provoke a change in unfolding pathway under force is intriguing. A dramatically increased responsiveness to force induced by ubiquitination could be instrumental in promoting unfolding by the proteasome.

Investigating the mechanism of this force-induced difference could prove insightful. MD and HDX experiments have suggested that the mechanism of destabilization for mono-ubiquitinated barstarK60 relies on an entropic penalty imposed by a substantial loss of conformational flexibility in helix 3 of barstar upon ubiquitination (93). This is especially interesting in light of how mechanical extension tends to reduce the impact of chain entropy on unfolding. If this loss in entropy is the reason for differences in rates of mechanical unfolding, then rates of mechanical unfolding should be responsive to changing the size of the moiety attached at the site of ubiquitination. One relatively simple way to attempt this could be to measure the impact of a ubiquitin-binding protein on the rate of unfolding.

As a caveat, while unfolding under force offers a more comparable set of conditions for proteasomal engagement than unfolding by chemical denaturation, the sets of conditions are not perfectly comparable. Pulling velocities of force-ramp experiments are ~100-fold faster than the more slowly-moving ATPase motors of the proteasome (125), which results in a higher force required for unfolding via force-ramp than via the proteasome; constant force experiments, such as the force-jump experiments, do not probe via pulling velocity at all. Handle attachment requires tethering a substrate on both ends, while the eukaryotic proteasome exerts its unfoldase activity with only one end of the molecule constrained. Furthermore, the site and geometry of engagement may be variable, rather than proceeding from near the termini as in these experiments.

While I have prepared mono-ubiquitinated substrates for three different single-lysine variants of barstar—K2, K60, and K78—I have thus far only characterized one site of ubiquitination, K60. Given that the impact of ubiquitination is known to be site-specific, I am interested to find out how ubiquitin influences the unfolding pathways of these other variants. K78 in particular could be intriguing to characterize given the lack of destabilization upon ubiquitinated observed in bulk experiments. Moreover, the mechanisms of destabilization conferred by K2 and K60 are known to be distinct (92), raising the tantalizing possibility that their mechanical unfolding behaviors could be distinguishable.

3.5 Methods

3.5.1 Purification of barstar substrates

Barstar substrates were prepared as described previously (92). Mutations were introduced via site-directed mutagenesis using primers encoding the proper base pair changes (IDT). *Escherichia coli* BL21 Rosetta 2 (DE3) cells were transformed with plasmids bearing barstarK2, barstarK60, and barstarK78 constructs (pSV033, pSV022, and pSV034, respectively). Constructs consisted of an N-terminal His_6-tagged maltose binding protein (MBP) followed by a connecting linker containing a PPPY recognition sequence for E3 (Rsp5) recruitment, followed by a prescission cleavage site, followed by the N-terminal His_6-tagged barstar single-lysine variant with cysteines at positions 0 and 82 (C0 and C23A), plus all lysines besides the specified position mutated to arginine.

Single colonies were grown overnight by shaking at 37˚C. 50 mL of this culture was added to 2 L LB broth (Fisher) containing kanamycin and shaken at 37˚C. Cultures were induced when the OD_{600} was between 0.5 and 0.8 with 1 mM IPTG, and expression proceeded for 3 hours at 37°C. Cells were pelleted by centrifugation and resuspended in 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 20 mM imidazole, supplemented with 1X Halt protease inhibitor cocktail (Thermo) and benzonase (Novagen). Cells were lysed by sonication. Lysate was clarified by centrifugation at 15,000 rcf, 4°C, 30 minutes.

Clarified lysate was allowed to batch bind to HisPur Ni^{2+}-NTA resin (Thermo) washed with 50 mM Tris pH 7.5, 150 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, and eluted with 50 mM Tris pH 7.5, 150 mM NaCl, 500 mM imidazole and 0.5 mM TCEP. Eluate was concentrated in an Amicon spin concentrator (Millipore) and passed over a Superdex200 16/60 size-exclusion column (GE Healthcare Life Sciences) equilibrated in 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP. Peaks corresponding to the MBP-barstar fusion were collected and quantified by UV-Vis absorption at 280 nm. 10% glycerol was added to substrates prior to being flash frozen and stored at -80˚C.

Non-ubiquitinated substrates were produced by digestion with prescission protease overnight at 4˚C to cleave the MBP scaffold and linker. Protein was then concentrated and loaded onto an S75 16/60 size-exclusion column (GE) pre-equilibrated in 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP in order to separate MBP scaffold and precision protease from the desired product, which was single-lysine barstar with an N-terminal His_6 tag. 10% glycerol was added to substrates prior to being flash frozen and stored at -80˚C.

3.5.2 Reductive methylation of ubiquitin

Ubiquitin was reductively octadimethylated as described by Hershko *et al.* (126). Briefly, 10 mg of bovine ubiquitin (Thermo Scientific) was dissolved in 10 ml of 100 mM HEPES-NaOH buffer (pH 7.4) containing 6 M urea (as verified by refractive index). Formaldehyde and sodium cyanoborohydride were then added to final concentrations of 12 and 20 mM, respectively. The reaction was allowed to proceed at room temperature, stirring, for 24 hours. After 24 hours, fresh formaldehyde and sodium cyanoborohydride were added at the above concentrations, and the reaction was permitted to proceed for one further hour. Unreacted reagents were removed by dialysis in 3 kDa cutoff dialysis tubing over the course of 48 hours against six total changes of buffer (100 mM HEPES, pH 7.4).

Total methylation of all lysines (>99.9%) was confirmed by native MS. A molecular weight of 8789 Da (Figure 3.11) is consistent with the anticipated molecular weight of the octadimethylated product, given the parent ubiquitin molecular weight of 8565 Da. No native or partially methylated ubiquitin is detected in the sample.

This methylated ubiquitin was compared against purchased methylated ubiquitin (Thermo Scientific) for its performance in the *in vitro* preparation of ubiquitination substrates. Equivalent performance in mono-ubiquitination reactions was observed for ubiquitin methylated in this manner and purchased methylated ubiquitin.

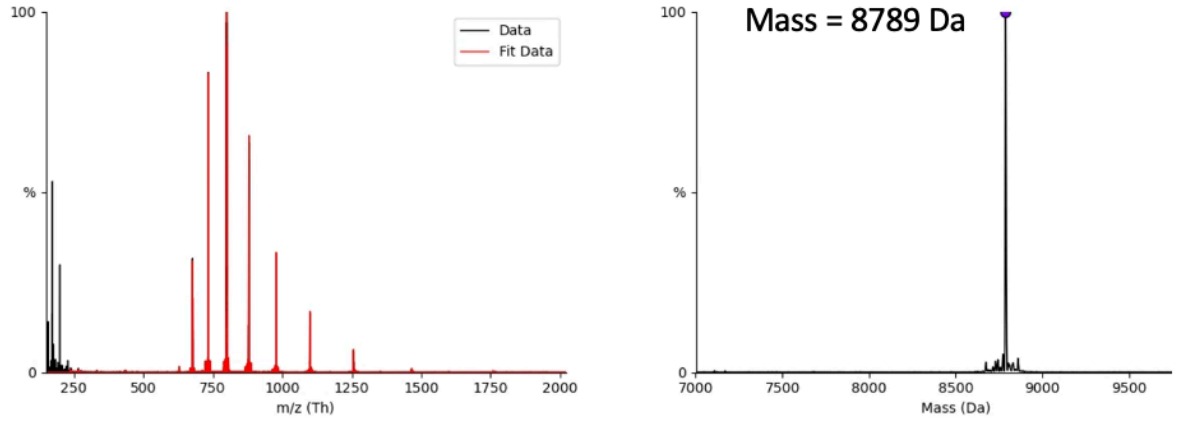


Figure 3.11: Mass spectrum and deconvoluted mass (Unidec) of octadimethylated ubiquitin, demonstrating complete conversion to desired product.

3.5.3 Purification of ubiquitination enzymes

Ubiquitination machinery components *M. musculus* mE1, *S. cerevisiae* Ubc4, and *S. cerevisiae* Rsp5 were purified as described previously (86, 92). *E. coli* BL21 Rosetta 2 (DE3) pLysS cells were transformed with pAM235 (mE1) or pAM236 (Ubc4) or pAM237 (Rsp5). Single colonies were grown overnight by shaking at 37˚C. 50 mL of this culture was added to 6 L LB broth (Fisher) containing kanamycin and shaken at 37˚C until OD_{600} was between 0.5 and 0.8. Cultures were induced with 1 mM IPTG, and expression took place overnight at 18°C. Cells were pelleted by centrifugation and resuspended in 50 mM HEPES pH 7.6, 250 mM NaCl supplemented with protease inhibitors (pepstatin A, aprotonin, PMSF, and leupeptin), benzonase, and lysozyme (2 mg/mL). Cells were lysed by sonication. Lysate was clarified by centrifugation at 15,000 rcf, 4°C, for 30 minutes and passed through a 0.2 µm filter.

Clarified lysate was allowed to batch bind to HisPur Ni^{2+}-NTA resin (Thermo) washed with 50 mM HEPES pH 7.6, 250 mM NaCl for one hour at 4°C. Resin was washed in a gravity flow column with 50 mL of 50 mM HEPES pH 7.6, 250 mM NaCl, 20 mM imidazole, and eluted with 50 mM HEPES pH 7.6, 250 mM NaCl, 500 mM imidazole. Eluate was concentrated in an Amicon spin concentrator (Millipore) and subsequently passed over a Superdex200 16/60 size-exclusion column (GE) equilibrated in 20 mM HEPES pH 7.6, 100 mM NaCl, 10% glycerol. Peaks corresponding to the MBP-barstar fusion were collected and quantified by UV-Vis absorption at 280 nm and flash frozen for storage at -80˚C.

3.5.4 Preparation of homogenous mono-ubiquitinated substrate proteins

Substrate proteins and ubiquitination enzymes were prepared as described above.

Ubiquitination reactions were set up as previously described (92), in ubiquitination reaction buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5 mM MgCl2, and 5% glycerol) in 20 µL aliquots as follows: 2.5 µM Uba1 (E1), 2.5 µM Ubc4 (E2), 5 µM Rsp5 (E3), 20 µM substrate, 500 µM methylated ubiquitin, and 5 mM ATP.

These aliquots were incubated in strip tubes in a thermocycler for 3 hours at 25°C. For a typical reaction, 48 individual 20 µL reactions were performed. Reactions were then pooled. To cleave the MBP scaffold, prescission protease was added and cleavage was permitted to proceed overnight at 4°C.

His-tagged ubiquitination machinery and ubiquitinated His-MBP substrate scaffold was removed by size exclusion chromatography. The reaction was concentrated to <500 µL and loaded onto an S75i 10/300 size exclusion column (GE) pre-equilibrated with 25 mM Tris pH 7.5, 150 mM NaCl, and 0.5 mM TCEP. The peak corresponding to the mono-ubiquitinated substrate was collected, concentrated, and quantified by UV-Vis absorption at 280 nm before flash freezing to store at -80°C.

3.5.5 Preparation and attachment of DNA handles to protein substrates

This is a modification of a protocol by Robert Sosa (direct communication).

Since long pieces of nucleic acid could not be synthesized directly, I created double-stranded DNA ready for direct attachment to the shorter handles using PCR. Forward primers were purchased from IDT functionalized with either 5'digoxigenin or 5'biotin, while reverse primers contained a BsaI cleavage site in order to produce 4 bp overhangs complementary to the annealed oligos following incubation with BsaI (NEB). Full-size handles were 558 bp long and amplified from the pGEMX plasmid.

4 bp overhangs are in bold, the BsaI recognition motif is underlined.

Oligo-1: 5-/phos/**AGGA**GTGTCTTCATATGATAGGAATAGGCGTAAC

Oligo-2: 5-/phos/**CAAC**GTGTCTTCATATGATAGGAATAGGCGTAAC

NH-Oligo (reverse): CACAGAAGTATACTATCCTTATCCGCATTG-NH(5AmMC6)

Design for full-size handle 1: forward strand = 5/Digoxigenin/XXXX…XGGTCTCC-3, reverse strand = 3-XXXX…XCCAGAGG**TCCT**-5

Design for full-size handle 2: forward strand = 5-/Biot/XXXX…XGGTCTCC-3, reverse strand = 3-XXXX…XCCAGAGG**GTTG**-5

Upon receipt, oligos were dissolved at a concentration of 400 µM in 10 mM Tris, 10 mM EDTA, 50 mM, NaCl, pH 7.4 and kept stored at -20˚C in 20 µL aliquots. Per handles reaction, 40 µL of each oligo (Oligo-1 and Oligo-2) was each mixed separately with 40 µL of NH-oligo. Anneal each set of oligos to form a dsDNA oligo with a 4 bp overhang by incubating at 95˚C for 5 min, then decreasing the temperature to 4˚C at a rate of 0.1˚C/s. Do not use a heated lid. Theoretically, these oligos may be combined, but I have found it more effective to functionalize and purify them separately to ensure precise stoichiometry at the final step. Micro Bio-Spin™ P-6 Gel Columns (BioRad) were used to buffer exchange the oligos into water, following the manufacturer's instructions. It is important to remove Tris in this step because GMBS is reactive with amine groups.

GMBS should be kept stored dessicated at 4˚C. It should be kept anhydrous and has low solubility in water. A fresh 40 mM stock in DMSO was prepared just before combining reaction components. NH-oligos were functionalized by combining reagents according to the following table, and then vortexing overnight (~16 hours) at room temperature.

Component	Initial [ ]	Final [ ]	Vol (µL)
NH-dsOligo	200 uM	133 uM	40
$NaCO_3$ pH 8.0	400 mM	33.3 mM	5
GMBS	24 mM	2 mM (24X NH-dsOligo)	5
DMSO	100	~17%	10

Each set of oligos was diluted to 500 µL using 10 µM Tris pH 8.5 (EB buffer, Qiagen) + 10% DMSO, prepared fresh. Oligos were concentrated to 100 µL in a 3K Amicon spin concentrator (Millipore) at 4˚C, 14,000 rpm. This was repeated five times total to wash away excess GMBS, then washed five more times using EB buffer alone. Oligos were concentrated to ~50 µL, quantified by Nanodrop, and kept cold. Aim for a final concentration of ~30 µM each oligo.

His_6-tagged protein engineered with two cysteines was reduced by incubation with 50 mM TCEP overnight at room temperature. Buffer exchange was performed using 1x CS buffer (100 mM sodium phosphate, 150 mM NaCl, pH 7.2, 0.003% tween-20) and 7K Zeba spin desalting columns (Thermo) according to manufacturer instructions. Tween-20 was always added fresh. It's theoretically not necessary to completely remove the reducing agent. Once the protein has been desalted, it must be added immediately to other reagents. Crosslinking was performed by combining reagents according to the following table:

Component	Initial [ ]	Final [ ]	Vol (µL)
CS-buffer	10X	1X	12
SH-Protein-SH	50 uM	1X	10
GMBS-dsOligo-1	30 uM	3X	50
GMBS-dsOligo-2	30 uM	3X	50

Reaction was shaken for 20 hours at 16˚C, 500 rpm. Varying proportion of oligo:protein ratio may improve yield on a case-by-case basis.

Finally, the final product was enriched by nickel purification. 300 µL of HisPur Ni^{2+}-NTA resin (Thermo) was equilibrated by washing five times with 500 µL of W buffer (25 mM Tris-HCl pH 7.6 and 200 mM KCl) by resuspension in a 1.5 mL tube and centrifuging at 4500 rpm. Beads were resuspended by addition of crosslinked protein-oligos and 400 µL W buffer. Beads were rotated for one hour at 4˚C, then collected by centrifugation through a 0.2 µm filter for ~1 min at 1500 rpm at 4˚C, and flowthrough was discarded. Beads were washed three times with 500 µL W buffer, then resuspended with 500 µL 25 mM Tris-HCl pH 7.6, 200 mM KCl, 1 M imidazole, rotated at 4˚C, 30 min, and eluant was collected. Protein-DNA chimera was desalted into 50 mM Tris pH 7.0, 250 mM NaCl using 7K Zeba spin desalting columns, and protein concentration was quanitifed by Nanodrop A_{280}.

3.5.6 Optical trap experimental setup

Streptavidin-coated beads (Pierce) were passivated by vortexing with 2 mg/mL BSA (NEB) for 10 minutes. Anti-digoxigenin beads were prepared by washing 1 mL of Spherotech Protein G Polystyrene Particles (3.18 micron), 0.5% w/v, cat # PGP-30-05, with 1 mL PBS pH 7.4. Beads were spun down and resuspended in 1 ml of 100 mM sodium phosphate pH 8.5, 100 mM NaCl. 60 µL of sheep polyclonal digoxigenin antibody (Roche 1333 089) and 30 µL of DMP (Pierce) were added, and beads were tumbled at room temperature for 60 min. Next, beads were spun down and resuspended in 1 mL of 2 M Tris (base) and vortexed for 2 hours at minimal speed to quench the crosslinking reaction. Finally, beads were diluted threefold with PBS pH 7.0, then pellet and resuspend the beads in 500 µL PBS pH 7.0 five times (4000 rpm, 2 min). Ensure the pH of the final product is neutral to avoid hydrolysis. Beads were stored at 4˚C.

On the day of experiments, full-length handles were ligated to oligo-protein chimeras. An example incubation is as follows: 1 µL each full-length handle (50 nM in water) with 2 µL oligo-protein chimera (125 nM) plus 1 µL T4 ligase (NEB), 1 µL 10x T4 ligase buffer (NEB), 4 µL water, incubated at 16˚C for 1 hour. Resulting ligations were incubated (1-10 µL, amount and dilution dependent on sample performance) with 1 µL anti-digoxigenin beads and 10 µL tweezers buffer (100 mM Tris pH 7.0, 250 mM NaCl) at room temperature for 20-40 minutes. Beads were then diluted in 1 mL of tweezers buffer before being injected into instrument fluidics.

Tweezers were calibrated according to a Stokes law velocity versus drag force calculation using 3.0 micron-sized sized beads (Pierce). Force-ramps were performed at a pulling rate of 100 nm/s from 3 to 65 pN. Force-jumps were performed using force feedback, with a folding force of 3 pN and unfolding forces between 6 and 20 pN. Sampling rate was 4000 Hz.

3.5.7 Data analysis for kinetics and for force jumps

Data were analyzed using a MATLAB suite of in-house analysis programs (Jesse Dill) and by using Python. Force ramps were extracted, and rips were identified based on force and extension change. Force clamp analysis was used to split data regions containing individual jump trajectories. A t-test was used to identify statistically significant changes in extension; each extension change was manually recorded to verify rip identity. Distinguishable intermediates were counted. Dwell times were recorded based on the first transition. Dwell times were pooled, and probability distribution functions were constructed using Python such that the bin size was less than 20% of the mean lifetime of the distribution.

References

1. K. Tunyasuvunakool, *et al.*, Highly accurate protein structure prediction for the human proteome. *Nature* **596** (2021).
2. K. A. Dill, J. L. Maccallum, The Protein-Folding Problem, 50 Years On (May 16, 2018).
3. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science (80-.).* **181** (1973).
4. C. Levinthal, How to fold graciously. *Mössbauer Spectrosc. Biol. Syst. Proc.* **24** (1969).
5. L. S. Itzhaki, G. D. Rose, Folding and binding: lingering questions, emerging answers. *Curr. Opin. Struct. Biol.* (2012) https:/doi.org/10.1016/j.sbi.2012.01.001.
6. B. Ma, C. J. Tsai, T. Haliloğlu, R. Nussinov, Dynamic allostery: Linkers are not merely flexible. *Structure* **19** (2011).
7. J. L. Harman, *et al.*, Evolution avoids a pathological stabilizing interaction in the immune protein S100A9. *Proc. Natl. Acad. Sci. U. S. A.* **119** (2022).
8. K. Röder, D. J. Wales, Evolved Minimal Frustration in Multifunctional Biomolecules. *J. Phys. Chem. B* **122** (2018).
9. G. Kramer, D. Boehringer, N. Ban, B. Bukau, The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* (2009) https:/doi.org/10.1038/nsmb.1614.
10. A. Ramanathan, P. K. Agarwal, Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. *PLoS Biol.* (2011) https:/doi.org/10.1371/journal.pbio.1001193.
11. L. L. Porter, L. L. Looger, Extant fold-switching proteins are widespread. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5968–5973 (2018).
12. F. Chiti, C. M. Dobson, Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* **86** (2017).
13. T. N. Shamsi, T. Athar, R. Parveen, S. Fatima, A review on protein misfolding, aggregation and strategies to prevent related ailments. *Int. J. Biol. Macromol.* **105** (2017).
14. T. R. Jahn, S. E. Radford, Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469** (2008).
15. C. Lee, M. P. Schwartz, S. Prakash, M. Iwakura, A. Matouschek, ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol. Cell* **7** (2001).
16. F. U. Hartl, M. Hayer-Hartl, Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16** (2009).
17. E. Braselmann, J. L. Chaney, P. L. Clark, Folding the proteome. *Trends Biochem. Sci.* **38** (2013).
18. K. W. Roskamp, C. N. Paulson, W. D. Brubaker, R. W. Martin, Function and Aggregation in Structural Eye Lens Crystallins. *Acc. Chem. Res.* **53**, 863–874 (2020).

19. H. Zhao, *et al.*, The molecular refractive function of lens γ-crystallins. **411**, 680–699 (2011).
20. H. Bloemendal, *et al.*, Ageing and vision: Structure, stability and function of lens crystallins. *Prog. Biophys. Mol. Biol.* **86** (2004).
21. S. Bassnett, On the mechanism of organelle degradation in the vertebrate lens (2009).
22. K. A. Reynolds, R. N. Mclaughlin, R. Ranganathan, Hotspots for allosteric regulation on protein surfaces https:/doi.org/10.1016/j.cell.2011.10.049 (January 29, 2019).
23. D. Pascolini, S. P. Mariotti, Global estimates of visual impairment: 2010. *Br. J. Ophthalmol.* **96** (2012).
24. J. A. Carver, J. A. Aquilina, P. G. Cooper, G. A. Williams, R. J. W. Truscott, α-Crystallin: molecular chaperone and protein surfactant. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **1204** (1994).
25. M. Delaye, A. Tardieu, Short-range order of crystallin proteins accounts for eye lens transparency. *Nature* **302**, 415–417 (1983).
26. N. E. N. E. Robinson, *et al.*, Quantitative measurement of young human eye lens crystallins by direct injection Fourier transform ion cyclotron resonance mass spectrometry. **12**, 704–711 (2006).
27. A. Basak, *et al.*, High-resolution X-ray crystal structures of human γD crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract. *J. Mol. Biol.* **328** (2003).
28. Z. Xia, Z. Yang, R. Zhou, et al., UV-radiation induced disruption of dry-cavities in human γD-crystallin results in decreased stability and faster unfolding. *Sci. Rep.* **3**, 1560 (2013).
29. S. L. Flaugh, I. A. Mills, J. King, Glutamine deamidation destabilizes human γD-crystallin and lowers the kinetic barrier to unfolding. *J. Biol. Chem.* **281** (2006).
30. E. Serebryany, *et al.*, Aggregation of Trp > Glu point mutants of human gamma-D crystallin provides a model for hereditary or UV-induced cataract. *Protein Sci.* **25**, 1115–28 (2016).
31. J. F. Hejtmancik, Congenital cataracts and their molecular genetics. *Semin. Cell Dev. Biol.* **19**, 134–149 (2008).
32. D. A. Stephan, E. Gillanders, M. J. Brownstein, et al., Progressive juvenile-onset punctate cataracts caused by mutation of the γD-crystallin gene. *Proc. Natl. Acad. Sci. USA* **96**, 1008–1012 (1999).
33. E. Nandrot, C. Slingsby, L. Hilal, et al., Gamma-D crystallin gene (CRYGD) mutation causes autosomal dominant congenital cerulean cataracts. *J. Med. Genet.* **40**, 262–267 (2003).
34. P. Evans, K. Wyatt, C. Slingsby, et al., The P23T cataract mutation causes loss of solubility of folded γD-crystallin. *J. Mol. Biol.* **343**, 435–444 (2004).
35. B. Wang, C. Yu, L. Xie, et al., A novel CRYGD mutation (p.Trp43Arg) causing autosomal dominant congenital cataract in a Chinese family. *Hum. Mutat.* **32**, E1939–E1947 (2011).
36. E. Héon, M. Priston, F. L. Munier, et al., The γ-crystallins and human cataracts: a

puzzle made clearer. *Am. J. Hum. Genet.* **65**, 1261–1267 (1999).
37. W. Zhang, H. C. Cai, Y. B. Yan, et al., The congenital cataract-linked G61C mutation destabilizes γD-crystallin and promotes non-native aggregation. *PLoS One* **6**, e20564 (2011).
38. J. Graw, J. Löster, M. H. de Angelis, et al., V76D mutation in a conserved gD-crystallin region leads to dominant cataracts in mice. *Mamm. Genome* **13**, 452–455 (2002).
39. M. Roshan, P. H. Vijaya, K. Satyamoorthy, et al., A novel human CRYGD mutation in a juvenile autosomal dominant cataract. *Mol. Vis.* **16**, 887–896 (2010).
40. W. Xu, *et al.*, A novel cataract-causing mutation Ile82Met of γA crystallin trends to aggregate with unfolding intermediate. *Int. J. Biol. Macromol.* **211** (2022).
41. K. L. Moreau, J. King, Hydrophobic core mutations associated with cataract development in mice destabilize human gammaD-crystallin. *J. Biol. Chem.* **284**, 33285–33295 (2009).
42. F. Ji, *et al.*, The human W42R γD-crystallin mutant structure provides a link between congenital and age-related cataracts. *J. Biol. Chem.* **288**, 99–109 (2013).
43. K. L. Moreau, J. King, Hydrophobic core mutations associated with cataract development in mice destabilize human γD-crystallin. *J. Biol. Chem.* **284**, 33285–33295 (2009).
44. I. A. Mills, S. L. Flaugh, M. S. Kosinski-Collins, J. A. King, Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γD-crystallin and γS-crystallin. *Protein Sci.* **16**, 2427–2444 (2007).
45. S. L. Flaugh, M. S. Kosinski-Collins, J. King, Contributions of hydrophobic domain interface interactions to the folding and stability of human D-crystallin. *Protein Sci.* **14**, 569–581 (2005).
46. M. S. Kosinski-Collins, In vitro unfolding, refolding, and polymerization of human gammaD crystallin, a protein involved in cataract formation. *Protein Sci.* **12**, 480–490 (2003).
47. M. S. Kosinski-Collins, S. L. Flaugh, J. King, Probing folding and fluorescence quenching in human gammaD crystallin Greek key domains using triple tryptophan mutant proteins. *Protein Sci.* **13**, 2223–2235 (2004).
48. P. Das, J. A. J. A. J. A. King, R. Zhou, Aggregation of γ-crystallins associated with human cataracts via domain swapping at the C-terminal β-strands. **108**, 10514–10519 (2011).
49. I. A. Mills-Henry, S. L. Thol, M. S. Kosinski-Collins, E. Serebryany, J. A. King, Kinetic Stability of Long-Lived Human Lens γ-Crystallins and Their Isolated Double Greek Key Domains. *Biophys. J.* **117**, 269–280 (2019).
50. M. J. Whitley, *et al.*, A Combined NMR and SAXS Analysis of the Partially Folded Cataract-Associated V75D γD-Crystallin. *Biophys. J.* **112**, 1135–1146 (2017).
51. C. N. Pace, K. L. Shaw, Linear extrapolation method of analyzing solvent denaturation curves. *Proteins Struct. Funct. Genet.* **41** (2000).
52. S. W. Englander, T. R. Sosnick, J. J. Englander, L. Mayne, Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6** (1996).
53. 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** (2019).
54. S. L. Flaugh, M. S. Kosinski-Collins, J. King, Interdomain side-chain interactions in human γD crystallin influencing folding and stability. *Protein Sci.* **14**, 2030–2043 (2005).
55. M. Petersen, D. Barrick, Analysis of Tandem Repeat Protein Folding Using Nearest-Neighbor Models. *Annu. Rev. Biophys.* **50**, 245–265 (2021).
56. K. Mandal, S. K. Bose, B. Chakrabarti, R. J. Siezen, Structure and stability of γ-crystallins. I. Spectroscopic evaluation of secondary and tertiary structure in solution. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **832** (1985).
57. J. Chen, S. L. Flaugh, P. R. Callis, J. King, Mechanism of the highly efficient quenching of tryptophan fluorescence in human γD-crystallin. *Biochemistry* **45**, 11552–11563 (2006).
58. K. L. Moreau, J. A. King, Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends Mol. Med.* **18** (2012).
59. T. O. Street, N. Courtemanche, D. Barrick, Protein Folding and Stability Using Denaturants. *Methods Cell Biol.* (2008) https:/doi.org/10.1016/S0091-679X(07)84011-8.
60. D. Barrick, R. L. Baldwin, Three-State Analysis of Sperm Whale Apomyoglobin Folding. *Biochemistry* **32**, 3790–3796 (1993).
61. T. Kajander, A. L. Cortajarena, E. R. G. Main, S. G. J. Mochrie, L. Regan, A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* (2005) https:/doi.org/10.1021/ja0524494.
62. D. E. Makarov, A theoretical model for the mechanical unfolding of repeat proteins. *Biophys. J.* **96** (2009).
63. J. K. Myers, C. Nick Pace, J. Martin Scholtz, Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* (1995) https:/doi.org/10.1002/pro.5560041020.
64. D. D. Weis, T. E. Wales, J. R. Engen, M. Hotchko, L. F. Ten Eyck, Identification and Characterization of EX1 Kinetics in H/D Exchange Mass Spectrometry by Peak Width Analysis. *J. Am. Soc. Mass Spectrom.* **17**, 1498–1509 (2006).
65. Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, Primary structure effects on peptide group hydrogen exchange. *Proteins Struct. Funct. Bioinforma.* **17** (1993).
66. Yu-Zhu Zhang, “Protein and peptide structure and interactions studied by hydrogen exchange and NMR,” University of Pennsylvania, Philadelphia, Pennsylvania. (1995).
67. P. Aravind, S. K. Suman, A. Mishra, Y. Sharma, R. Sankaranarayanan, Three-dimensional domain swapping in nitrollin, a single-domain betagamma-crystallin from Nitrosospira multiformis, controls protein conformation and stability but not dimerization. *J. Mol. Biol.* **385**, 163–77 (2009).
68. S. Lee, *et al.*, A single destabilizing mutation (F9S) promotes concerted unfolding of an entire globular domain in γS-crystallin. *J. Mol. Biol.* **399** (2010).
69. J. C. Boatz, M. J. Whitley, M. Li, A. M. Gronenborn, P. C. A. Van Der Wel, Cataract-associated P23T γD-crystallin retains a native-like fold in amorphous-

looking aggregates formed at physiological pH. *Nat. Commun.* **8** (2017).
70. E. Serebryany, J. A. King, The βγ-crystallins: Native state stability and pathways to aggregation. *Prog. Biophys. Mol. Biol.* **115** (2014).
71. E. Serebryany, *et al.*, An Internal Disulfide Locks a Misfolded Aggregation-prone Intermediate in Cataract-linked Mutants of Human D-Crystallin * (2016) https:/doi.org/10.1074/jbc.M116.735977.
72. E. Serebryany, J. A. King, Wild-type human γd-crystallin promotes aggregation of its oxidation-mimicking misfolding-prone W42Q mutant. *J. Biol. Chem.* **290**, 11491–11503 (2015).
73. S. Garcia-Manyes, *et al.*, Single-molecule force spectroscopy predicts a misfolded, domain-swapped conformation in human γD-crystallin protein. *J. Biol. Chem.* **291**, 4226–4235 (2016).
74. A. Horwich, Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J. Clin. Invest.* **110** (2002).
75. L. Acosta-Sampson, J. King, Partially Folded Aggregation Intermediates of Human γD-, γC-, and γS-Crystallin Are Recognized and Bound by Human αB-Crystallin Chaperone. *J. Mol. Biol.* **401** (2010).
76. A. J. Guseman, *et al.*, Assessing the Structures and Interactions of γD-Crystallin Deamidation Variants. *Structure* **29**, 284-291.e3 (2021).
77. M. Newville, A. Ingargiola, T. Stensitzki, D. B. Allen, Lmfit: Non-Linear Least-Square Minimization and Curve-Fitting for Python (2014) https:/doi.org/10.5281/ZENODO.11813.
78. J. D. Marold, *et al.*, A collection of programs for one-dimensional Ising analysis of linear repeat proteins with point substitutions. *Protein Sci.* **30**, 168–186 (2021).
79. G. R. Masson, *et al.*, Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16** (2019).
80. R. Yau, M. Rape, The increasing complexity of the ubiquitin code. *Nat. Cell Biol.* **18** (2016).
81. F. Ohtake, Branched ubiquitin code: From basic biology to targeted protein degradation. *J. Biochem.* **171** (2022).
82. A. Werner, *et al.*, Cell-fate determination by ubiquitin-dependent regulation of translation. *Nature* **525** (2015).
83. G. L. Grice, J. A. Nathan, The recognition of ubiquitinated proteins by the proteasome. *Cell. Mol. Life Sci.* **73** (2016).
84. B. Coll-Martínez, B. Crosas, How the 26S proteasome degrades ubiquitinated proteins in the cell. *Biomolecules* **9** (2019).
85. E. R. Greene, K. C. Dong, A. Martin, Understanding the 26S proteasome molecular machine from a structural and conformational dynamics perspective. *Curr. Opin. Struct. Biol.* **61** (2020).
86. J. A. M. Bard, *et al.*, Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **87** (2018).
87. J. Callis, The Ubiquitination Machinery of the Ubiquitin System. *Arab. B.* (2014) https:/doi.org/10.1199/tab.0174.

88. L. Bedford, J. Lowe, L. R. Dick, R. J. Mayer, J. E. Brownell, Ubiquitin-like protein conjugation and the ubiquiting-proteasome system as drug targets. *Nat. Rev. Drug Discov.* **10** (2011).
89. P. Kulkarni, *et al.*, Structural metamorphism and polymorphism in proteins on the brink of thermodynamic stability. *Protein Sci.* (2018) https:/doi.org/10.1002/pro.3458 (September 8, 2018).
90. T. Tomita, A. Matouschek, Substrate selection by the proteasome through initiation regions. *Protein Sci.* **28** (2019).
91. T. Hagai, A. Azia, Á. Tóth-Petróczy, Y. Levy, Intrinsic disorder in ubiquitination substrates. *J. Mol. Biol.* (2011) https:/doi.org/10.1016/j.jmb.2011.07.024.
92. E. C. Carroll, E. R. Greene, A. Martin, S. Marqusee, Site-specific ubiquitination affects protein energetics and proteasomal degradation. *Nat. Chem. Biol.* **16**, 866–875 (2020).
93. E. C. Carroll, *et al.*, Mechanistic basis for ubiquitin modulation of a protein energy landscape. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).
94. P. Roca-Cusachs, V. Conte, X. Trepat, Quantifying forces in cell biology. *Nat. Cell Biol.* (2017) https:/doi.org/10.1038/ncb3564.
95. G. A. Collins, A. L. Goldberg, The Logic of the 26S Proteasome. *Cell* **169** (2017).
96. G. Stirnemann, S. Kang, R. Zhou, B. J. Berne, How force unfolding differs from chemical denaturation. *Proc. Natl. Acad. Sci.* (2014) https:/doi.org/10.1073/pnas.1400752111.
97. E. J. Guinn, B. Jagannathan, S. Marqusee, Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat. Commun.* **6**, 1–8 (2015).
98. J. Schönfelder, D. De Sancho, R. Perez-Jimenez, The Power of Force: Insights into the Protein Folding Process Using Single-Molecule Force Spectroscopy. *J. Mol. Biol.* **428** (2016).
99. A. Ashkin, J. M. Dziedzic, S. Chu, Observation of a single-beam gradient-force optical trap for dielectric particles in air. *Opt. Lett.* **22** (1986).
100. C. Cecconi, E. A. Shank, F. W. Dahlquist, S. Marqusee, C. Bustamante, Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur. Biophys. J.* (2008) https:/doi.org/10.1007/s00249-007-0247-y.
101. S. B. Smith, L. Finzi, C. Bustamante, Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads. *Science (80-.).* **258** (1992).
102. C. Bustamante, Z. Bryant, S. B. Smith, Ten years of tension: Single-molecule DNA mechanics. *Nature* **421** (2003).
103. K. C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* (2008) https:/doi.org/10.1038/nmeth.1218.
104. R. A. Maillard, *et al.*, ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* (2011) https:/doi.org/10.1016/j.cell.2011.04.010.
105. K. N. Swatek, D. Komander, Ubiquitin modifications. *Cell Res.* **26** (2016).

106. D. Komander, M. Rape, The Ubiquitin Code. *https://doi.org/10.1146/annurev-biochem-060310-170328* **81**, 203–229 (2012).
107. R. Khurana, J. B. Udgaonkar, Equilibrium unfolding studies of barstar: Evidence for an alternative conformation which resembles a molten globule. *Biochemistry* **33**, 106–115 (1994).
108. F. N. Zaidi, U. Nath, J. B. Udgaonkar, Multiple intermediates and transition states during protein unfolding. *Nat. Struct. Biol. 1997 412* **4**, 1016–1024 (1997).
109. B. Nölting, *et al.*, The folding pathway of a protein at high resolution from microseconds to seconds. *Proc. Natl. Acad. Sci. U. S. A.* **94** (1997).
110. R. B. Best, *et al.*, Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J. Mol. Biol.* (2003) https:/doi.org/10.1016/S0022-2836(03)00618-1.
111. Z. Zhou, *et al.*, Genetically encoded short peptide tags for orthogonal labelling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* (2007) https:/doi.org/10.1021/cb700054k.
112. C. P. Guimaraes, *et al.*, Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8** (2013).
113. C. S. Theile, *et al.*, Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8** (2013).
114. S. Vijay-kumar, C. E. Bugg, W. J. Cook, Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194** (1987).
115. M. J. Lubienski, M. Bycroft, S. M. V. Freund, A. R. Fersht, Three-Dimensional Solution Structure and 13C Assignments of Barstar Using Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **33** (1994).
116. Y. Seol, J. Li, P. C. Nelson, T. T. Perkins, M. D. Betterton, Elasticity of Short DNA Molecules: Theory and Experiment for Contour Lengths of 0.6-7 mm https:/doi.org/10.1529/biophysj.107.112995.
117. L. Bongini, V. Lombardi, P. Bianco, The transition mechanism of DNA overstretching: A microscopic view using molecular dynamics. *J. R. Soc. Interface* **11** (2014).
118. C. Jarzynski, Nonequilibrium fluctuations of a single biomolecule. *Lect. Notes Phys.* **711** (2007).
119. G. E. Crooks, Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **90** (1998).
120. R. Petrosyan, Improved approximations for some polymer extension models. *Rheol. Acta* **56** (2017).
121. B. Jagannathan, S. Marqusee, Protein folding and unfolding under force. *Biopolymers* **99** (2013).
122. G. I. Bell, Models for the specific adhesion of cells to cells. *Science (80-.).* **200** (1978).
123. I. Tinoco, P. T. X. Li, C. Bustamante, Determination of thermodynamics and kinetics of RNA reactions by force. *Q. Rev. Biophys.* **39**, 325–360 (2006).
124. M. Hinczewski, C. M. Gebhardt, M. Rief, D. Thirumalai, From mechanical folding

trajectories to intrinsic energy landscapes of biopolymers. *Proc. Natl. Acad. Sci. U. S. A.* **110** (2013).
125. J. A. M. Bard, C. Bashore, K. C. Dong, A. Martin, The 26S Proteasome Utilizes a Kinetic Gateway to Prioritize Substrate Degradation. *Cell* **177** (2019).
126. A. Hershko, H. Heller, Occurrence of a polyubiquitin structure in ubiquitin-protein conjugates. *Biochem. Biophys. Res. Commun.* **128**, 1079–1086 (1985).