2021

Quantifying the Release of Protein Substrates from AAA+ ATPase ClpX by Single Molecule Total Internal Reflection Fluorescence Microscopy

Xiao Wang

Quantifying the Release of Protein Substrates from AAA+ ATPase ClpX by Single Molecule Total Internal Reflection Fluorescence Microscopy

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Xiao Wang

June 2021

Quantifying the Release of Protein Substrates from AAA+ ATPase ClpX by Single Molecule Total Internal Reflection Fluorescence Microscopy

Xiao Wang, Ph.D.
The Rockefeller University 2021

The safe disposal of proteins is an essential process for maintaining proteostasis in cells. AAA+ proteases, such as proteasomes, play a major role in selective degradation of proteins. Degradation by AAA+ proteases typically requires the substrate to be physically unfolded before proteolysis. The efficiency of the unfolding process is hence a critical parameter for determining the turnover rate of the substrate. However, the mechanism by which force is utilized to unfold the protein substrate is not fully characterized. One parameter regulating the efficiency of the unfolding process is the rate by which a substrate is prematurely released before unfolded. Importantly, a propensity to be prematurely released by AAA+ proteases has been proposed to be an important mechanism allowing some substrates to escape proteolysis, but factors that influence the substrate release rate have not been well characterized. In this thesis, we investigate the parameters that influence the premature release of the substrate using the AAA+ ATPase ClpXP. We present a new assay based on Total-Internal-Reflection Fluorescence (TIRF) microscopy that measures the lifetime of the ClpX-substrate complex. We demonstrate that the technique has the potential to identify factors that affect the mean lifetime of the ClpX-substrate complex. Using this method, we show that the substrate release rate can be reduced by simply increasing the length of the unstructured region where ClpX grips the substrate. Next, we demonstrate that reducing the frequency of the ClpX mechanical cycle prolongs the lifetime of ClpX-substrate interaction. We also offer evidence that the association of the protease particle ClpP to the AAA+ particle stabilizes the interaction between ClpX and the protein substrate, even when the substrate does not directly interact with ClpP. Finally, we show that the glycine-alanine repeats, which are associated with proteins that cannot be fully degraded by the proteasomes, do not increase the release rate of the substrate from ClpX. The results indicate that the force ClpX exerts on the substrate can lead to the release of the substrate when it cannot be unfolded. The results support the hypothesis that the transient loss of grip between ClpX and its substrate could be a frequent side-effect of the substrate unfolding process, but ClpX is often able to recover its grip before the substrate dissociates.

I dedicate this thesis to my parents, who sent me out to let me see the world

ACKNOWLEDGMENTS

My deepest gratitude goes to my two mentors that guide me through this course of study—my advisor Dr. Sanford Simon, and our collaborator and friend Dr. Philip Coffino. This work wouldn't be possible without Dr. Simon's support and encouragement during the entire period of my PhD, whose courage and dedication to trying to solve new problems will always inspire me. I am also grateful for Dr. Coffino's invaluable knowledge, patience and generosity, and the rigorous approach to science.

I would also like to acknowledge the Rockefeller University graduate program, the Dean's Office and all of its members for offering me this opportunity to undertake this study. I am thankful to my committee member, Dr. Hironori Funabiki and Dr. Michael O'Donnell, for their insightful feedbacks and kind encouragements to help me always on track.

I'd like to give special thanks to my teammate Dr. Bassem Shebl, for our daily scientific discussions on science. I am also deeply indebted the supports I received from my colleagues Dr. Marina Bleck, Dr. Danial Johnson and Dr. Joan Pulupa on single molecule microscopy.

I offer my thanks to the specialists of the university resource centers that I had the pleasure to work with: Dr. Carlos Rico and Dr. Carolina Alcaino.

I am grateful for being able to work with a fantastic group of colleagues, for their helps on all matters related to both science and life. I would like to specially thank Dr. Kate Bredbenner, Dr. Melissa Jarmel, Gabriella Spitz-Becker, and Marcello Im, who have always supported me when I needed the most.

Last, I want to thank my family: to my parents who have always encouraged me to pursue my dream, and to my wife, Wen, whose love and care keeps me strong.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAA+	ATPases Associated with diverse cellular Activities, which is the name of a protein family with many shared features
ATPγS	Adenosine 5'-[γ-thiol]triphosphate
ATP	Adenosine triphosphate
ADEPs	Acyldepsipeptides
a.a.	amino acid (used as a unit for measuring polypeptide length)
RKH loop	arginine-lysine-histidine loop of ClpX
ClpP	ATP-dependent Clp protease proteolytic subunit
ClpX	ATP-dependent Clp protease subunit X
BSA	bovine serum albumin
cpGFP	circular-permutated green fluorescent protein
cryo-EM	cryogenic electron microscopy
DDS	dichlorodimethylsilane
DHFR	dihydrofolate reductase
EF	enzyme-fragment complex
ES	enzyme-substrate complex
EDTA	ethylenediaminetetraacetic acid
FOV	field of view
ssrA	gene name short for "small stable RNA"
GRR	glycine rich region
GAr	glycine-alanine repeats
GAr15	glycine-alanine repeats of 15 amino acid long
IGF loop	isoleucine-glycine-phentolamine loop of ClpX
MTX	methotrexate
NADH	nicotinamide adenine dinucleotide
PEP	phosphoenolpyruvate
PolyQ	polyglutamine
PD buffer	protein Degradation Buffer
SR	serine-rich
SPR	surface plasmon resonance
ssrA tag	the degron tag added to carboxyl-terminus of protein substrates
ssrADD mutant	the ssrA with its two carboxyl-end alanine residues mutated to aspartic acid
TIRF	Total Internal Reflection Fluorescence
ZBD	Zinc-binding-domain

Chapter 1
The Current Understandings of ClpX and ClpXP

1.1 Introduction to ClpXP

Proteins require controlled synthesis and destruction to regulate their quantity and quality in response to the needs of the organism. One major pathway for regulating the destruction of proteins is by the AAA+ proteases. The prime example of the AAA+ protease is the eukaryotic 26S proteasomes, which has a general architecture that is shared across all branches of life (Pickart and Cohen, 2004). The AAA+ protease typically consists of two parts—the gated proteolytic barrel, the lumen of which contains the active sites of protein degradation, and the AAA+ ATPase motor sitting at the gate of the protease particle. This architecture reflects the two general steps of AAA+ protease mediated degradation, translocation-coupled unfolding and proteolytic cleavage. In this basic scheme, the ring-shaped hexameric AAA+ ATPase motor controls access to the inside of the protease particle by selectively translocating and unfolding tagged protein substrate through its narrow central pore into the protease particle. Once there, the already unfolded substrate is safely cleaved by the protease in isolation.

ClpXP is a bacterial proteasome-like complex. In many ways, ClpXP and the proteasome function analogously. In comparison to the 26S proteasome, however, ClpXP is much simpler in its subunit composition, thus representing one of the most basic forms of a proteasome-like complex. The complex is assembled from two subcomplexes, ClpX and ClpP. The AAA+ ATPase ClpX, is a homohexamer, while ClpP is a homo-tetradecamer, composed of two stacked rings, each with seven-fold symmetry. Together, the ClpXP complex degrades protein using only two types of building blocks, making this system an attractive model for understanding the AAA+ protease in general (Sauer and Baker, 2011).

Impeded protein unfolding by AAA+ proteases can lead to profound physiological consequences. On the one hand, interrupted degradation of individual proteins can be an important mechanism for generating normally required cellular proteins. For example, NF-κB1 is activated by partial degradation of its precursor p105 by the proteasome (Yu, Wan and Hua, 2009). On the other hand, EBNA1 protein found in Epstein-Barr virus impairs antigen presentation by proteasomes by escaping the AAA+ mediated unfolding (Zhang and Coffino, 2004). It is estimated that the 26S proteasome is responsible for 70% of cellular protein degradation (Voges et al., 1999). This raises the question of how a single protein complex processes an immense variety of substrates, and how some proteins are degraded but not others.

To answer these questions, it is essential to understand the basic biophysical process of substrate unfolding by AAA+ ATPase. Thanks to its simplicity of architecture, ClpXP has been used as an accessible system for modeling the biophysical process of proteasome-mediated degradation. Despite extensive prior biophysical and biochemical experimentation, the dynamics of ClpXP during the unfolding of substrate is not fully understood. In this thesis, I explore the question of substrate release. Specifically, what factors could determine that a substrate is released without degradation by the AAA+ protease. I show that for ClpXP, accidental release is potentially frequent but is mitigated by biophysical parameters like the length of an unstructured region of the substrate and the rate of ClpX ATP hydrolysis. To investigate the factors that affect the substrate release rate, I developed a novel single molecule microscopy assay to monitor the dissociation of unprocessed substrate proteins from ClpXP, or ClpX alone. I demonstrated the validity of this assay, and described dynamic interactions within and between substrates and the ClpX motor which determine the release process. This thesis identifies gaps in the current understanding of how the substrate is unfolded, where future experiments could be directed.

In Chapter 1, I provide an overview on the current understandings of ClpX, which has been drastically reshaped by several newly published studies on the cryo-EM structure of ClpXP complex. With new insights gained from the cryo-EM structures, I re-visit previous functional studies of ClpX and ClpXP and evaluate some previous conclusions that can be updated by the new structural insights.

In Section 1.2, I survey the new structural findings, and use these structures as an introduction for basic ClpX properties. In Section 1.3, I discuss previous biochemical studies on ClpX and ClpXP and how the evidence supports the newly published structures. In Section 1.4, I cover the current understandings of the substrate unfolding process. In Section 1.5, I conclude the chapter by introducing the key question of this study: what factors govern the release rate of unprocessed substrate.

1.2 The Structure of ClpXP

1.2.1 An overview of the structure of ClpXP

To understand the mechanism of substrate degradation, it is essential to know how ClpX uses energy, how ClpX binds its substrates, and what kinds of conformational changes allow ClpX to exert force. While structural biology may offer many direct answers to the above questions, it has been technically challenging to obtain ClpX structures. The first ClpX crystal structure was described in 2003 (Kim and Kim, 2003) from *Helicobacter pylori*, although unsuccessful attempts were made by the same group to obtain ClpX structures from various other bacterial species. In 2009, Glynn et al. obtained the *E. coli* ClpX crystal structures for the nucleotide-free and nucleotide-bound states (Glynn et al., 2009), and this study has been instrumental in informing subsequent functional studies on ClpX. It was not until 2019 and early 2020 that the cryo-EM structures for ClpXP became available (Gatsogiannis et al., 2019; Ripstein et al., 2020; Fei et al., 2020), where major differences between the crystal structures and the cryo-EM structures became apparent with regard to ClpX nucleotide binding states and arrangements of subunit conformations around the ring.

Each monomeric ClpX contains an N-terminal zinc-binding-domain (ZBD) and an AAA+ ATPase domain. The function of the ZBD is presumably for recognition of various degradation tags, and is not essential for the function of the ATPase domain (Martin et al., 2005). As a result, the ZBD was deleted from the ClpX crystal structures by Glynn et al. The AAA+ domain contains a large subdomain and a small subdomain; both are structurally rigid. A flexible hinge connects the large domain and the small domain. From the crystal structure it is clear that the hinge allows large shifts in the relative orientations of the large and the small domains.

ClpX monomers assemble into a hexameric ring by association between the small domain of one subunit and the large domain of the adjacent subunit upon nucleotide binding. When viewed from above the ring (top view defined as the side where substrate enters the central pore; also shown in Figure 1.2.1), subunits are named in alphabetical order from A to F, and join into a ring in the clockwise direction. ATP binds at the pocket formed between the large and small domain of each subunit, causing each hexamer to have six potential ATP binding sites, although not all of them necessarily have the same state of nucleotide occupancy. This is reflected both by the crystal structure and the cryo-EM structure, which I will discuss in greater detail in section 1.2.2.

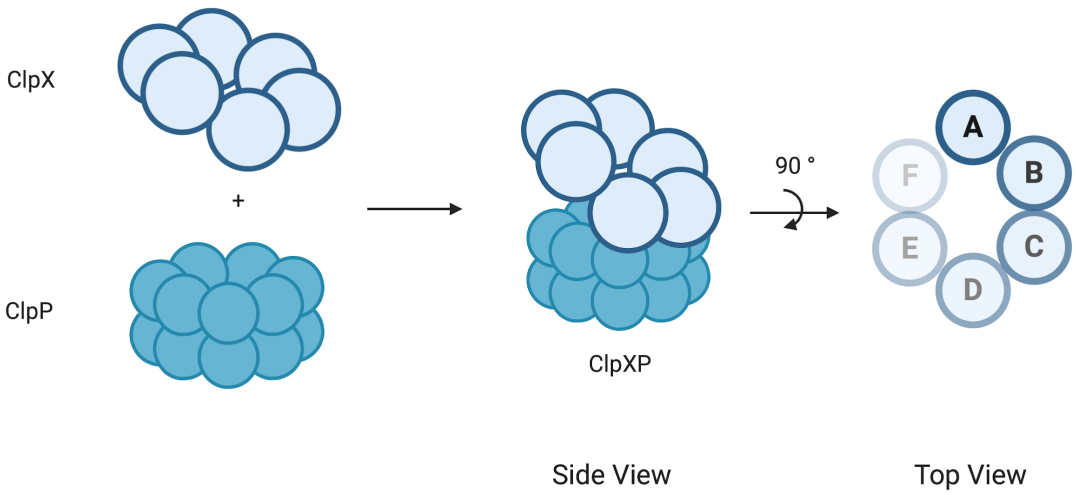


Figure 1.2.1 Definition of side views and top views used in this thesis.

Each ClpX subunit contains pore loops of different functions, and there are six copies for each type of loops in the hexamer. The RKH loop (named for the tripeptide Arg-Lys-His) extends above the ClpX ring and is implicated in initiating the engagement of the substrate. Pore 1 loops contain the conserved tyrosine that is presumably responsible for gripping the substrate and transmitting the pulling force from ClpX. Within each subunit, the Pore 2 loop sits below the Pore 1 loop, and is implicated in maintaining processivity. I will discuss the properties of the pore loops in Section 1.2.3.

Below the ClpX ring are the IGF loops (named for the tripeptide Ile-Gly-Phe), which are responsible for docking to ClpP. There was very little structural insight on how this binding occurs between ClpX and ClpP prior to the publications of the cryo-EM structures of ClpXP. The ClpP barrel has a seven-fold symmetry while ClpX has a six-fold symmetry, so the binding between the two will necessarily be slightly mismatched. The two cryo-EM structures have shown that the IGF loop is sufficiently long to accommodate the symmetry mismatch.

The general architecture of the ClpP has been well characterized by crystallography and electron microscopy (Wang et al., 1997; Ortega et al., 2000; Szyk and Maurizi, 2006; Bewley et al., 2006). In the absence of AAA+ particle, ClpP can only degrade short peptides (Woo et al., 1989), suggesting that free diffusion into the barrel by larger substrate is prevented. Crystal structures have identified that the gateway into the protease barrel is regulated by the N-terminal sequences of ClpP, the deletion of which allows efficient degradation of large unfolded proteins without needing help from an AAA+

unfoldase such as ClpX (Bewley et al., 2009). This indicates a gating mechanism at the ClpP barrel. It was speculated that IGF loop binding might change the conformation of the ClpP gate (Lee et al., 2010), allowing ClpX to translocate unfolded substrate into the ClpP barrel for degradation. From many biochemical analyses it is clear that binding of ClpP impacts ClpX activity, although without a cryo-EM structure of ClpX ring alone it is difficult to draw the conclusion of the structural impact. The structural changes of ClpP and its likely effect on ClpX confirmation will be discussed in section 1.2.4.; the effects of the ClpP binding on ClpX activity will be discussed in great detail in Section 1.3.4.

1.2.2 The hand-over-hand model describing the ClpX mechanical cycle

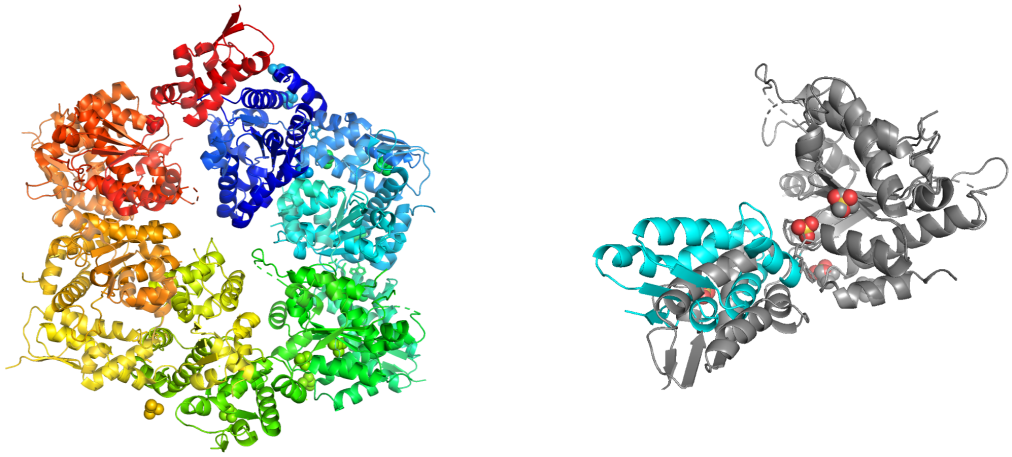


Figure 1.2.2 Left: top view of the crystal structure of hexameric ClpX, with 4 nucleotides bound. (Glynn et al., 2009; rendered from PDB ID: 3HWS using Pymol). Right: two conformations (referred to as L and U) of the subunits from the same structure, differed by the relative rotation of the large and small domains. The U conformation is highlighted in Cyan.

Two conformations of ClpX subunits (Figure 1.2.2) are reported by Glynn et al. in 2009, and the difference between them is the relative orientations of the large and small domains. One of the two conformations allows ATP binding, and is therefore referred to as “L” conformation which stands for ATP-Loadable; the other conformation is called “U”, which stands for ATP-Unloadable and excludes ATP binding, due to steric hindrance. For both nucleotide-free and nucleotide-bound structures, the subunits from A to F adopts an U-L-L-U-L-L arrangement, and all the interfaces between subunits (namely, A/B, B/C, C/D, D/F, F/A) are the same across the ring. This 4L-2U arrangement is interpreted to allow 4 nucleotides to bind per hexamer at any given time. In 2005, Hersch et al. used using

isothermal titration calorimetry to determine nucleotide binding stoichiometry. An ATPase-dead mutant of ClpX, which was able to bind but unable to hydrolyze ATP, was found to bind 3.4 ± 0.5 ATP (Hersch et al., 2005). In 2013, Stinson et al. utilized a novel FRET technique to distinguish between the L and U conformations, and observed that under saturating-concentration of nucleotides, ClpX adopted a 5L-1U conformation (Stinson et al., 2013).

Gatsogiannis et al. (2019) reported cryo-EM structures of ClpXP from *Listeria monocytogenes*, and observed that all subunits adopt a conformation similar to "L". Another novel feature of their structure is that they did not remove the ZBD, and observed a strange head-to-head dimerization between two hexametric ClpX via the ZBD. The significance of this dimerization is not clear.

Notably this structure revealed that the Pore 1 loops of ClpX adopts a spiral staircase arrangement, similar to that observed in proteasomes (de la Peña et al., 2018). In contrast, this spiral staircase arrangement is not observed in previously published ClpX crystal structures (Figure 1.2.3 shows a rendered cryo-EM structure of the spiral staircase).

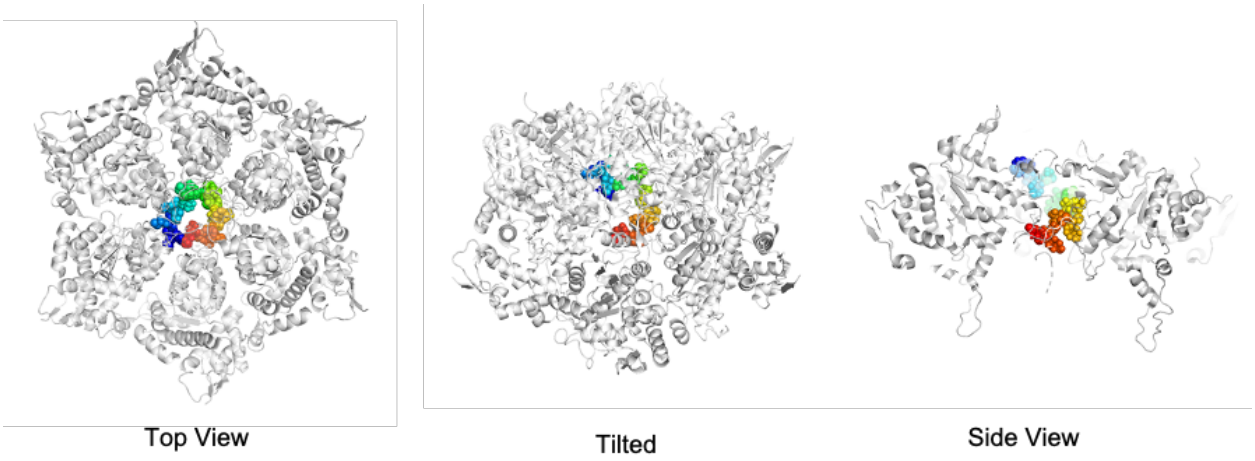


Figure 1.2.3 Top, tilted, and side views of the same Cryo-EM structure of the ClpX, Class 1, from Fei et al. (2020). The tyrosine residues in the Pore 1 loops are highlighted as colored spheres. (PDB ID: 6POS)

Gatsogiannis et al., however, did not resolve the nucleotide binding states of individual subunits. This problem was resolved by cryo-EM structures from Ripstein et al. (2020) and Fei et al. (2020). Once again, all subunits adopt the L-like conformation in all classes of structures from these two groups. In structures from Fei et al., the *E. coli* ClpXP has five ATPγS and one ADP bound. Ripstein et al. obtained their structures from *Neisseria meningitidis*; one conformation has 4 ATP and 2 ADP bound and the other conformation

has 5 ATP and 1 ADP bound. Notably, the spiral staircase arrangement was also confirmed by Ripstein et al. and Fei et al. It is unknown, however, whether nucleotide binding to all six subunits is a necessary condition for adopting the spiral staircase.

To simplify nomenclature, when viewing ClpXP from the side, the topmost subunit is named as A, and the spiral goes down from A to F (Figure 1.2.4), joined at the same type of interface observed in crystal structures. Analysis from Fei et al. showed that the relative orientation between the large and small domains within each subunit varies, highlighting the elasticity of the hinge region which accommodates the ring closure. ADP is found in either A subunit or F subunit, or in the structures from Ripstein et al., both A and F subunits. Considering the predicted direction of substrate movement, this finding implies that ATP hydrolysis might power a conformational change from F to A.

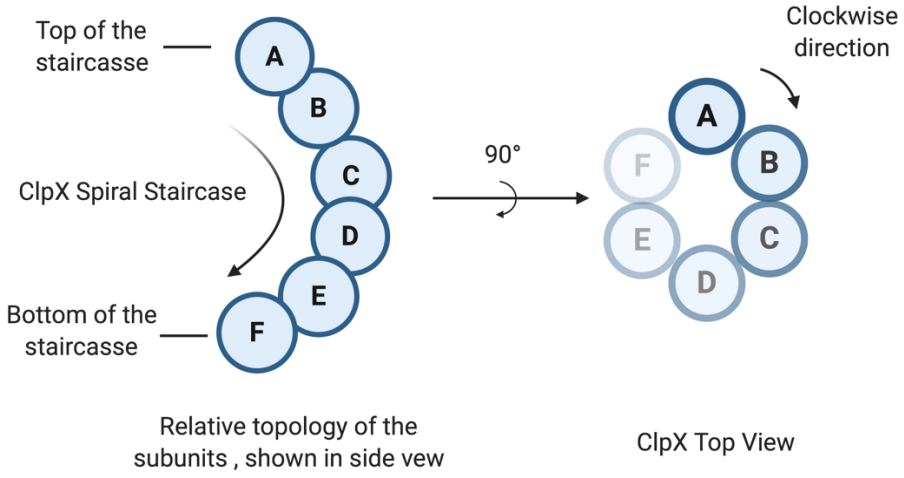


Figure 1.2.4
The naming convention and relative position of the subunits used in this thesis

Studies by Ripstein et al. and Fei et al. were able to resolve structural details of the individual ATPase sites in the ClpX ring. In general, ClpX has classic Walker motifs essential for interacting with phosphate groups and catalyzing ATP hydrolysis (Walker et al., 1982). In addition, several conserved AAA+ ATPase features were resolved in these structures, including the Arginine Finger and Sensor II. The Arginine Finger contributes to the hydrolysis of ATP in the adjacent subunit (in the counterclockwise direction when viewed from the top; for example, arginine finger from F contributes to hydrolysis of ATP in A) by coordinating the leaving γ-phosphate group in its neighbor's ATPase site. Sensor II senses the binding of the nucleotide by interacting with the phosphate groups. ClpX lacks a conserved Sensor I, which regulates the hydrolysis of ATP by coordinating a water

molecule that is needed for the reaction, and Fei et al. was able to identify a glutamine residue that functions like a Sensor I residue in close proximity to the γ-phosphate of the bound nucleotide. Similar to the Arginine Finger, Sensor I functions by contributing to the ATPase activity of the counterclockwise neighboring subunit. Considering the location of these motifs in the structure, it appears that F does not have an active ATPase pocket because the Sensor I and Arginine Finger of A are pulled away due to the slight rotation of the small domain of F subunit.

Before the publication of ClpX cryo-EM structures, it was assumed that the conformational change driven by ATP hydrolysis is one that shifts from L to U conformation, with four ATP bound per hexamer. The absence of U conformation in cryo-EM structures poses the question whether U conformation exists at all, or only transiently under physiological conditions, or only observed under crystal-packing conditions that favor low energy states. On the other hand, a large body of biochemistry studies indicate that ClpX translocates substrate with a step size that cannot be exactly explained by the relatively small conformational shift from that of F to that of A. I will discuss this topic in more detail in Section 1.3.2 and 1.4.2 when reviewing published optical tweezers experiments.

Assuming that a conformational change happens where the F subunit moves to the top of the spiral, this motion strongly agrees with the proposed model shared by several AAA+ unfoldases, such as Cdc48, YME1, and the proteasomal AAA particle (Ripstein et al., 2017; Puchades et al., 2017; de la Peña et al., 2018). In this general model, which is referred to as the “hand-over-hand” pulling model by these authors, the bottom subunit of the spiral disengages binding with the substrate, moves to the top, then re-engages the substrate. This motion of polypeptide translocation is reminiscent of the hand-over-hand motion when a person climbs a rope. Because the engagement with the substrate is mediated by pore loops, the hand-over-hand model will be restated in terms of tyrosine loops in Section 1.2.3 when their function is discussed.

However, the hand-over-hand mechanism implies a sequential order of ATPase firing, which in the case of ClpX contradicts mutational studies on ClpX ATPase, wherein ClpXP still supports substrate degradation even when up to three subunits are defective in ATPase activity (Martin et al., 2005**)**. This is in stark contrast with mutational studies on proteasomes, where ATPase mutation of each subunit leads to significant defect for substrate degradation *in vitro* (Kim et al., 2013). Therefore, one key area that future studies should address is to resolve the conflicts of these models.

1.2.3 The arrangement of pore loops and interactions with protein substrate

As mentioned above, there are three classes of pore loops in ClpX: the RKH loops (Sequence: VPPQGGRKHPQQEFL), the Pore 1 loops (sequence: GYVG), and the Pore 2 loops (sequence: PSITRDVSG). The RKH loops protrude above the ClpX ring, with its two positively charged RK residues essential for engaging the canonical bacterial C-terminal degradation tag, the ssrA tag (Martin et al., 2008b). The Pore 1 and Pore 2 loops populate the central pore and lie close to each other. The Pore 1 loop contains the essential tyrosine residue which has been repeatedly seen in other AAA+ ATPases and is crucial for the unfolding of substrate (the function of tyrosine loops will be discussed in great detail in Section 1.4). The Pore 2 loop does not have large amino acid residues, but the sequence contains charged and non-polar residues. For each ClpX subunit, the Pore 1 loop is located closer to the top of the ring than the Pore 2 loop, but due to the spiral staircase arrangement of the subunits, several Pore 2 loops interdigitate Pore 1 loops when going down the spiral. It has been proposed for 26S proteasomes that the Pore 1 loops are used to drive the substrate forward by steric interaction with the substrate side chains by the conserved tyrosine residue, while Pore 2 loops facilitate the engagement of the substrate by interacting with its backbone (de la Peña et al., 2018).

Because the tyrosine loop is most likely used for driving substrate translocation, the hand-over hand model is best illustrated by the motion of these tyrosine loops. As shown in the diagram in Figure 1.2.5, the tyrosine loop at the bottom disengages from the substrate, moves to the top, then reengages the substrate. The repeated cycle propels the substrate in a stepwise fashion.

However, while the hand-over-hand model shows that the tyrosine loop paddles the substrate in one direction, the biophysical nature of this interaction is not very well established. The contact between the pore loops and the substrate in most reported cryo-EM structures is not sufficiently resolved to infer whether the tyrosine loop interacts with the substrate by gripping the amino acid side chains or polypeptide backbone. In the proteasome cited study above, de la Peña et al. (2018) showed that the orientation of the tyrosine residues support their model that the interaction is mediated by side chains, but this conclusion would predict that side chain size would be the only limiting factor in the rate of substrate translocation or unfolding. Bell et al. (2019) conducted a very comprehensive scanning of the effect of diverse substrate side chains on substrate degradation activity using ClpX, and discovered that degradation efficiency was not only a function of size of the side chain but also its charge and position. In the specific case of ClpX structure by Fei et al., the bound substrate was presumed to be an adventitiously trapped endogenous substrate, of unknown and heterogeneous identity, and hence was modeled as a poly-alanine chain. Here, ClpX uses both tyrosine and the adjacent valine

in the Pore 1 loop to interact with the β-carbon of the modeled alanine residue, suggesting an interaction driven by hydrophobicity. In the structures from Ripstein et al., bound substrate was also modeled by a poly-alanine chain due, to poorly resolved density, where the Pore 1 loops were shown to interact with the substrate backbone. Considering the extensive diversity of sequences that can be processed by ClpX, it is likely that the pore loops support multiple modes of interaction, including interactions with both side chains and backbones.

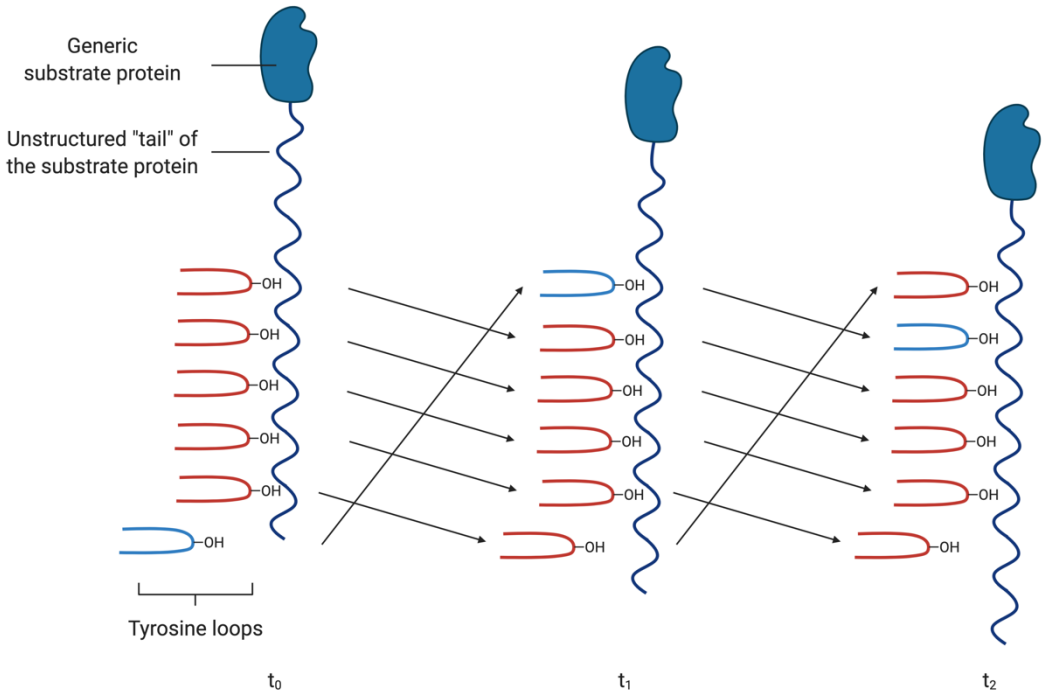


Figure 1.2.5. The hand-over-hand model. A “spiral staircase” of tyrosine loops is represented by the hydroxyl group on the tyrosine. The relative positions of the loops in the diagram correspond to their relative order within the spiral staircase. t0, t1, and t2 indicate the changing timepoints. Arrows indicate the movement of the tyrosine loops after each duty cycle.

Another problem that is not resolved by the hand-over-hand model is what happens when substrate stalls at ClpX. When ClpX unfolds a substrate, it may make multiple thwarted attempts to unfold, persisting over many ATPase cycles. In this process, ClpX is still able to hydrolyze ATP (Kenniston et al., 2003). If ATP hydrolysis is mechanically couple to conformational changes, the dynamic of this process prior to successful substrate unfolding is not satisfactorily explained by the current structures. Just focusing on the interaction between substrate and ClpX pore loops, there are at least two hypothetical scenarios that explain how ClpX could hydrolyze ATP without unfolding a substrate. In

the first case, the pore loops never disengage the substrates, such that the conformational change would not move past a transition stage, and only deform the substrate which springs back to its original structure and allows ClpX to fall back to the previous state. In the second case, the pore loops disengage the substrate briefly, such that the substrate slips back a step, allowing ClpX to perform a conformational change without moving forward. This scenario is akin to a rope climber, who goes through all the proper climbing motions, but gets nowhere because the rope is too slippery for his grasp. It is possible that the two scenarios are not mutually exclusive, but to test these cases, it is important to understand the biochemical nature of the interaction between the substrate and ClpX, which require functional studies in addition to structural analysis. In Section 1.4, I will focus on the question of slippery substrates, which is one of the key questions for understanding this process.

1.2.4 Interaction between ClpX and ClpP

It has been well established that association with ClpP affects various activities of ClpX, (This topic is revisited in Section 1.3.3). This suggests a potential allosteric regulation of ClpX activity upon binding to ClpP. Because there is no cryo-EM structure of ClpX in the absence of ClpP to use as a reference (and it is difficult to interpret the results from previous crystallography studies for the reason mentioned in Section 1.2.2), we can only infer the effects of ClpX-ClpP binding by analyzing the details of the interaction itself.

ClpX binding to ClpP is mediated by the IGF loops from each subunit, as determined in early genetic and biochemical studies (Kim et al., 2001; Joshi et al., 2004), but how the 6 IGF loops of ClpX match the 7 hydrophobic clefts on the top of ClpP had been a puzzle. The cryo-EM structures show that this is not a problem, because the length and flexibility of the IGF loops can adequately compensate for the symmetry mismatch (Gatsogiannis et al., 2019; Ripstein et al., 2020; Fei et al., 2020). Gatsogiannis et al. showed that the single empty hydrophobic cleft on ClpP sits right below subunit C; in structures from Ripstein et al. and Fei et al. the empty cleft is below subunit E. This difference may suggest that the binding between IGF loops and the hydrophobic clefts is dynamic, and can accommodate a large spectrum of structural changes of ClpX.

It was shown biochemically that progressively adding ClpP to a solution of ClpX could reduce the ATPase hydrolysis rate of ClpX (Joshi et al., 2004). Because IGF loops mediate most of the contacts between ClpX and ClpP, it is reasonable to hypothesize that changes in ClpX ATPase activities could be mediated by changes in dynamics of the IGF loops. Since the IGF loop is directly connected to the helical region that contains Sensor I and Arginine Finger, tension in the IGF loop might affect the ATPase activity of the

adjacent subunit. Vice versa, ATP binding may also limit how much the IGF loop can be extended, and thus may impact the affinity between ClpX and ClpP ring. When ClpX fires ATPases around its ring, it is likely that individual IGF loops need to disengage and re-engage ClpP to accommodate conformational changes (Amor et al., 2016), which may also have an effect on the kinetics of ClpX ATPase activity.

In the absence of ClpX, structures of ClpP showed that the openings of the barrel are very narrow, which prevents folded proteins from entering the protease barrel by accident in the absence of the AAA+ particle (Bewley et al., 2006). These openings are gated by the N-terminal sequences of ClpP monomers, which can be widened when the small-molecule drug acyldepsipeptides (ADEPs) bind the hydrophobic clefts (Kirstein et al., 2009). While Gatsogiannis et al. reported that IGF binding on the same hydrophobic clefts on ClpP does not induce gate opening in ClpP, Ripstein et al. and Fei et al. reported the opposite.

To reconcile these two observations, I propose that the protein substrate could play a key role in gate opening. Beside IGF loops, Gatsogiannis et al. reported contacts between Pore 2 loops and the N-terminal loops of ClpP in the absence of protein substrates, and they hypothesized that this interaction might regulate ClpP gate. Upon substrate engagement at the pore, Pore 2 loops are redirected to the substrate (Ripstein et al., 2020; Fei et al., 2020), which in turn could free the N-terminal sequences of the ClpP and allow its gate opening.

Another potential interaction between ClpX and ClpP is not resolved by the cited cryo-EM structural studies. When ClpXP unfolds and degrades substrates, the unfolded protein substrate stretches from the top of the ClpX to the active site of ClpP via the ClpP gate. It is possible that the presence of such substrate, with its simultaneous interaction with both ClpX and ClpP, further stabilizes the complex and brings about additional conformational changes in both ClpX and ClpP conformations.

1.3 The ATPase cycle of ClpXP

For ClpX, the conformational change that mechanically unfolds and translocates the protein substrate is dependent on hydrolysis of ATP. The properties of the ClpX ATPase cycle, together with its coupling to the mechanical cycle, reveal the mechanism of interaction with protein substrates. This is especially informative due to the limited information from structural studies.

Remarkably, the ATPase activity of ClpX is highly adaptive. ClpX hydrolyzes ATP at around 100-200 ATP per hexamer per minute (Burton et al., 2003; Joshi et al., 2004). Binding to ClpP significantly reduces ATP hydrolysis rate (Kim et al., 2001; Joshi et al., 2004). In the presence of protein substrates, however, ClpXP hydrolyzes ATP at a faster rate, and the rate is correlated with the stability of the substrate (Kenniston et al., 2003; Martin et al., 2008a). These observations reflect two possibilities. First, the binding to ClpP allosterically changes the ATPase kinetics of ClpX, as discussed in Section 1.2.4. Second, ATP hydrolysis is closely linked with the rate of conformational changes, such that a stable substrate may impede the movement of ClpX and slow the ATP hydrolysis rate, while an unstructured substrate facilitates conformational excursions, thus promoting ATP hydrolysis.

In Section 1.3, I review the past biochemical studies on ClpX ATPase activities in light of the recent cryo-EM structures. In 1.3.1, I discuss the coupling between ATPase cycle and the mechanical cycle, with the focus on ATPase activity. Section 1.3.2 covers the ATP cycle of ClpX, inferred from optical tweezers studies that measure mechanical cycles. Section 1.3.3 covers studies on ATPase mutants, and the implications of the ATPase firing model of ClpX. In 1.3.4, I briefly discuss the effect of ClpP binding on ClpX activity. In Section 1.4, I will revisit some of these topics but with the focus on interaction with the substrate.

1.3.1 The mechanical cycle of ClpX is coupled with the ATPase cycle

There are two levels of ATPase cycles in ClpX. On the individual subunit level, the ATPase undergoes ATP binding, hydrolysis, and release of ADP and phosphate; on the level of the hexameric complex, the ATPase cycle of each subunit needs to be integrated within the ring. Following certain step(s) within the hydrolysis cycle, ClpX performs the conformational change that drives the substrate through the central pore. This coupling between the mechanical cycle and the ATPase cycle of ClpX is referred to as mechanochemical coupling.

The available tools for studying mechanochemical coupling have been limited. So far, mechanical cycles can only be monitored by either measuring the rate of protein substrate degradation, or by measuring ClpX movements at a single molecule over a stretched substrate using optical tweezers. Both methods rely on changes in the substrate as a read-out for ClpX conformational changes, and are therefore not suitable to study situations where ClpX conformational change might fail to move or unfold substrates. Nonetheless, these studies have identified two types of perturbations of ClpX structure that disrupt substrate degradations without stopping ATPase cycle.

The first decoupling perturbation is on the hinge region (Figure 1.3.1) between the large and small domains of ClpX subunits. Glynn et al. showed that a single amino acid insertion into the hinge region keeps ClpX hexamers at basal ATPase activity but abolishes more than 90% of its unfoldase activity (Glynn et al., 2012). Bell et al. further investigated the hinge region using both linker deletions and linker extensions, and discovered that while hinge deletion leads to defects in ClpX ring closure, linker extension with 12 glycine residues not only causes defects in substrate degradation, but also increases both $\mathbf{V_{max}}$ and $\mathbf{K_m}$ for ATP; at the same time both perturbations decrease the Hill constant of ATPase hydrolysis, with the linker extension having a stronger effect (Bell et al., 2018), suggesting a reduction of cooperativity between individual ATPase subunits.

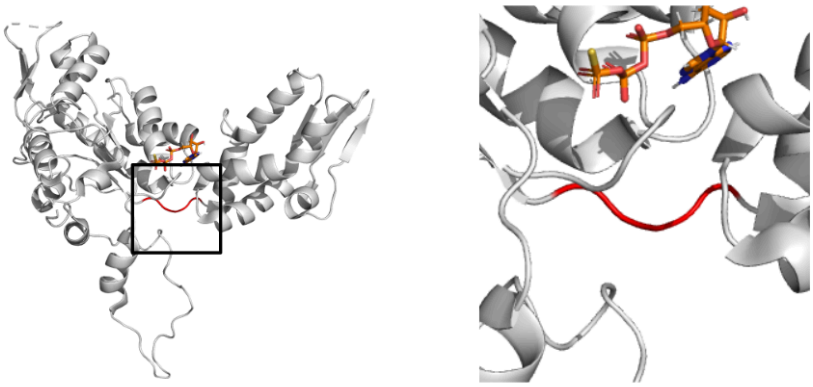


Figure 1.3.1. The hinge region of the ClpX subunit, highlighted in red. Left: a single ClpX subunit viewed from the side, with its IGF loop at the bottom. Right: zoomed in view of the hinge region marked. It can be seen that the bound nucleotide is in close proximity to the hinge (PDB: 6POS)

The second decoupling perturbation is the locking of conformations. Stinson et al. (2013) described a strategy of locking the relative orientation of the large and small domains of a single ClpX subunit in the hexameric ring using a disulfide bond; this perturbation led to a massive increase in ATPase hydrolysis rate and significant defects in protein

degradation activity in ClpXP (Stinson et al., 2013). In the same study, locking two subunits in L form further increased ATPase hydrolysis rate while repressing substrate degradation.

Findings from these two types of perturbations lead to two propositions. First, the hinge must have optimal length. To understand the importance of the hinge region, a modeling study done by Hwang and Lang showed that if all six subunits were to adopt an L conformation, ClpX should naturally form a spiral. Ring closure under these circumstances would compress the spiral, leading to mechanical tensions in the hinge region (Hwang and Lang, 2013). Therefore, the hinge should hypothetically be short enough to store tension upon ring closure, which allows communication between substrates via mechanical strains. At the same time, the hinge cannot be too short, or it make ring closure unfavorable or impossible. This hypothesis predicts that loss of ring closure in wild-type ClpX, which may happen due to causes such as substrate-mediated mechanical impacts, might also lead to disruption of mechanochemical coupling.

The second proposition is that stopping the ClpX conformational change can disrupt the mechanochemical coupling. This points to the possibility that ClpX might be able to perform a futile ATP cycle without changing conformation when the energy barrier to conformational excursions is too high.

Both propositions need to be considered in the extreme case where ClpX tries to unfold a substrate that could potentially deform the ClpX ring or stall ClpX conformational change altogether. Interestingly, both types of decoupling perturbations result in an increase of ATPase activity. This could imply that coupling with the mechanical cycle keeps ATPase activity in check. In other words, the mechanical counterforce received by ClpX during substrate unfolding may feedback to the ATPase cycle via the coupling mechanism. This could explain why ClpX reduces ATPase hydrolysis rate during substrate unfolding (Kenniston et al., 2003), which I will discuss in more detail in Section 1.4.2.

1.3.2 The power-stroke model of conformational change, and the ATPase cycle, revealed by single molecule studies

Because of mechanochemical coupling, the mechanical cycle can be perturbed by manipulating the ATPase cycle. With the help of optical tweezers and ATPase inhibitors targeting specific steps in the ATPase cycle, the properties of the ClpX ATPase cycle has been extensively characterized at single molecule resolution. In such experiments, ClpX and protein substrates are separately conjugated to two beads, and the two beads are

linked by the interaction between substrate and ClpX. An unfolded substrate can be stretched apart by the optical tweezers into a straight polypeptide; translocation of the substrate into ClpXP will therefore shorten the distance between beads. The change of distance thus reflects ClpX activity. The system can also measure and adjust the opposing forces on the beads to assess how much force ClpX can generate, and how its action is altered by an external opposing counterforce.

In 2011, two studies of optical tweezers measurements on ClpXP were published by Rodrigo Maillard et al. and Marie-Eve Aubin-Tam et al. (Maillard et al., 2011; Aubin-Tam et al., 2011) Both studies characterized basic biophysical parameters of ClpXP, including the maximum force generated by ClpXP (up to 20 pN), the fundamental step size of pulling on the substrate (1 nm), the inverse relationship between external pulling force and the translocation speed, and the effect of the presence or absence of ClpP. I will revisit most of these topics relating to the substrates and ClpP in Section 1.4. In this section, I will focus on findings related to the ATPase cycle.

One important parameter that attracts attention is the power output of ClpX. Maillard et al. calculated that if each 1 nm step taken by ClpX results from a single hydrolysis of ATP, the maximum output for ClpX is close to 5 kT, which is higher than the thermodynamic background, thus implying a power-stroke model. In this power-stroke motor, the energy from ATP hydrolysis is directly harnessed to power a large conformational change that drives the 1 nm step, which is equivalent to 4 a.a. per step when the substrate is extended by a 13 pN force. Similar observations were made by Aubin-Tam et al., where they reported a 5-7 a.a. step size, with at least 5 kT work per step, when pulling at about 13 pN. However, in the cryo-EM structure, the hypothetical conformational change from F-like to A-like predicts a 2 a.a. step size (Ripstein et al., 2020; Fei et al., 2020). This implies that currently available structures do not necessarily cover all possible conformations.

The observation of 1 to 4 nm steps was followed up in a subsequent study by Maya Sen et al. published in 2013, where they described a "burst-and-dwell" mechanical cycle (Sen et al., 2013). It is assumed that 1 nm represents the single fundamental step ClpX can take. According to their analysis, in the burst phase ClpX is able to take from one to four (unresolved) steps almost simultaneously, which is then followed by a dwell-phase of 350 ± 20 ms.

The same observation on variable numbers of steps was later corroborated by Cordova et al. in 2014 (Cordova et al., 2014). Importantly, Cordova et al. reported that there was no sequential correlation for numbers of steps taken over multiple bursts. Strikingly, when four out of six subunits are deactivated by ATPase mutation, ClpX is still able to perform

4 nm steps. This suggests a model of coordinated ATPase cycles between subunits, and a stochastic model for steps taken. At the same time this result argues against the assumption that ClpX has a tight mechanochemical coupling. It also poses the question whether ClpX uses a power-stroke model, because the energy calculation supporting a power-stroke model was based on the assumption that one ATP hydrolysis leads to 1 nm step. Consequently, I cannot confidently draw the conclusion that the 1-4 nm steps are results of 1-4 corresponding ATPase hydrolysis events, and I use the term "power-stroke" with some reservation. This problem also makes it harder to use optical tweezers data to help understand the behavior of the ClpX ATPase cycle. (I will discuss this result again in Section 1.3.3, when contrasting it with other ATPase mutants.)

The 1 nm unit step size does not agree with the hand-over-hand model inferred from cryo-EM studies (Ripstein et al., 2020; Fei et al., 2020). It should be noted that polypeptides have a very short persistence length, making it difficult to translate lengths measured in nanometers to equivalent lengths measured in number of amino acid residues. Maillard et al. reported that under 13 pN load, a 1 nm step is equivalent to 4 a.a. However, the hand-over-hand model predicts that the fundamental size is 2 a.a. Currently, this contradiction has not been resolved.

To dissect the ATPase cycle further, Sen et al. showed that phosphate release is the force generating step by excluding other steps in the ATPase cycle; they further showed that the size of the burst is dependent on ATP concentration, but not duration of the dwell-phase (Sen et al., 2013).

To analyze the coordination between the subunits, Sen et al. inhibited individual ATPases by titrating ATPγS. ClpX takes 20-40 times longer to hydrolyze ATPγS compared to ATP (Sen et al., 2013), and Sen et al. showed that addition of ATPγS induces long pauses in ClpX during threading of the substrate. Based on the concentration of ATPγS used, they estimated that loading of at least three ATPγS molecules per ClpX are needed to induce a pause.

In a study published in 2016 by Piere Rodriguez-Aliaga et al., a similar experiment was done in which case two ATPγS were shown to be sufficient to stall ClpX (Rodriguez-Aliaga 2016). Regardless of the discrepancy in numbers, a common ground between the two experiments is that binding of a single ATPγS is not enough to induce a pause in the hexamer ring.

Rodriguez-Aliaga et al. also showed that binding of ATPγS can suppress the step sizes during burst phases. These experiments support a model where the number of steps

taken by ClpX is dependent on the amount of ATP hydrolyzed during the burst phase. However, this conclusion is at odds with that of Cordova et al., and there is no further reason to reject either conclusions.

Together, the evidence from optical tweezers studies supports the following model of ClpX ATPase cycle. ClpX can stochastically take from 1 nm to 4 nm steps, with a weak correlation between step size and ATP concentration. After concerted ATP hydrolysis in multiple subunits, rapid and irreversible release of phosphate by each subunit leads to a conformational change, which may result in a step forward along the substrate polypeptide, and multiple such steps in rapid succession constitute a burst. Although each additional ATP hydrolysis contributes to a larger step size, the number of ATPs hydrolyzed cannot fully explain the resulting size of the step. The presence of one or two ATPγS molecules does not fully prevent the burst, but the inhibited subunit will not participate in the generation of a power stroke.

Apparently, this model of the ATPase cycle is incomplete, and the uncertainties are unsatisfying. This model has no information on the exact subunits in the hexamer that are involved in the hydrolysis, or the exact conformational changes resulting from the hydrolysis, or what determines the length of the dwell phase.

With the evidence from recently published cryo-EM studies, some of these uncertainties can be rationalized, but new problems arise. The burst phase can be easily rationalized as consecutive subunits firing in a hand-over-hand manner. However, this leaves the problem of individual step size unanswered.

1.3.3 A model for the ClpX ATPase cycle around the ring

To study the subunits involved in a burst phase, one monumental tool is the pseudo-oligomer of ClpX. In a study published in 2005, Andreas Martin et al. designed constructs with multiple ClpX domains covalently linked by L20 linkers, the length of which allows formation of oligomers only in one way (Martin et al., 2005). Martin et al. deleted the ClpX N-terminal ZBD domain to improve solubility, and created covalently linked pseudo-dimers, -trimers, and hexamers of ClpX. These can self-assemble into functional hexamers, for example a dimer of linked trimers. Remarkably and importantly, these diverse assemblages were all similarly functional in degradation assays, and similar compared to a native wild-type ClpX complex constituted of monomers. With this strategy, Martin et al. was able to control the number and position of ATPase mutant subunits within the assembled hexameric ring.

Two informative types of ATPase mutations of ClpX were thus studies. The E185Q mutation in the Walker B motif abolishes ATP hydrolysis but still allows ATP binding (Hersch et al., 2005). The R370K is a mutation of the Sensor II that abolishes ATP hydrolysis, but also at the same time blocks ClpX allosteric changes associated with ATP binding, as reflected by reduced affinity of ClpX to ClpP (Joshi et al., 2004). ATP binding mutation has been reported before (Stinson et al., 2013) but it is less common, and has not been tested in the same comprehensive fashion as by Martin et al. Therefore, I will focus on E185Q and R370K mutants in this section. Following the naming convention of Martin et al., I refer to wild type subunits as W, and the two mutant forms as E (for E185Q), and R (for R370K) respectively. A wild-type pseudo-hexamer is written as W-W-W-W-W-W; similarly, a wild-type hexamer assembled from two pseudo-trimers is referred to as W-W-W/W-W-W.

Martin et al. showed that the ATPase activities of W-W-W-W-W-R, W-W-W-W-R-R, and W-W-W-R-R-R are similar. When degrading denatured substrates, all three retain about 60% of wild-type ATPase activity rate, albeit having different numbers of wild-type subunits. In contrast, the substrate degradation rate is progressively reduced with incremental addition of R-class subunits (retaining 70%, 56%, and 44% of wild-type V_{max}, respectively); the same trend is observed for degradation of native substrates. These results suggest that ClpX is very good at bypassing ATPase defects within the ring. Martin et al. interpreted this as evidence that ClpX does not fit a strict sequential ATPase hydrolysis model, nor a concerted hydrolysis model where all six subunits need to hydrolyze ATP at the same time.

The same study reported similar ATPase and degradation activities for W-W-W-W-R-R and W-W-R/W-W-R. However, hexamers with three R-class subunits, W-R/W-R/W-R have a slower ATPase rate and degradation rate than W-W-W-R-R-R. This suggests that ClpX prefers to have wild-type subunits placed next to each other for efficient substrate degradation when three R-class subunits are present. As I will discuss shortly, this could either be due to the R-class mutation having an inhibitory effect to the adjacent wild-type subunit, or that the power-cycle prefers to have consecutive wild-type subunits lined up together.

Because R-class subunits affect binding to ClpP, to further reduce the number of wild-type subunits per hexamer Martin et al. instead used E-class subunits. In the most extreme case, the W-E-R-E-E-R hexamer is still able to degrade denatured substrates at a very slow but detectable rate (at 3% of wild-type activity). Therefore, a single subunit is sufficient to power substrate degradation when the energy barrier is low, suggesting that each power-stroke can be provided by the hydrolysis of one subunit.

When comparing W-W-E/W-W-E to W-W-R/W-W-R, the former has a more severe drop in ATPase activity, suggesting that E and R-class mutations are not equivalent (even though both E and R-class mutations completely abolishes the ATPase activity of the mutated subunit).

Interestingly, W-R-E/W-R-E has 66% lower ATPase activity than R-W-E/R-W-E, although both have the same number of W, E, R class subunits. Without knowing the exact mechanism, it appears that the E class subunit inhibits the ATPase activity of the adjacent subunit in the clockwise direction (when viewed from the top), while the R-class mutation does not have the similar inhibitory effect, or is at least more permissive than the E-class. This evidence suggests that ClpX has an allosteric mechanism, potentially related to Sensor II, that allows inter-subunit communication, such that ATP hydrolysis is promoted after the counterclockwise neighboring subunit hydrolyzes ATP. One prediction of this mechanism is that the ATPases firing is sequential in the clockwise direction.

Finally, let us consider W-W-R-E-E-R. While this construct has 16% lower ATPase activity compared to R-W-E/R-W-E, its degradation activity is 82% lower. This suggests that having four consecutive ATPase mutant subunits is disruptive to the mechanical cycle. Combining this result with the observation where ClpX prefers wild-type subunits to be arranged in tandem, it points to a mechanism whereby up to four consecutive subunits are involved in a duty cycle.

Stinson et al. in 2013 characterized a new type of ATPase mutant for ClpX, which is referred to as the VI class mutations. The VI mutant was created by a double mutation V78A/I79A, which reduces ATP binding affinity. This was utilized to study conformational changes of ClpX with a novel FRET technique, by which they were able to detect the ATP binding state of the labeled subunits. Comparing the following three constructs, W-VI-W*/W-VI-W*, W-VI*-W/W-VI*-W, and W*-VI-W/W*-VI-W, where * denotes the location of the ATP binding FRET probe, Stinson et al. showed that ATP binding is suppressed in the W subunit counterclockwise to VI, but not so for the W subunit clockwise to VI. This mechanism works in the opposite direction compared to the ATP hydrolysis suppression by the E185Q mutations. One hypothesis from this observation is that this mechanism allows consecutive subunits to be orderly loaded with ATP in the counterclockwise direction.

It is tempting to combine the conclusions from ATPase mutant studies with those derived from optical tweezers experiments. However, there is a big hurdle that needs to be mentioned. Cordova et al. in 2014 reported optical tweezers measurements using R-W-E/R-W-E (Cordova et al., 2014). This mutant ClpX has more frequent and longer pauses

between steps, an effect similar to the consequence of ATPγS binding. Cordova et al. argued that the dwells between bursts are also prolonged, although at this point it is no longer easy to distinguish dwells from pauses. One striking result is that this mutant is able to take 4 nm steps even with only two wild-type subunits. As discussed in 1.3.2, ClpX was observed to take 1 to 4 nm steps, with 1 nm steps assumed to be the fundamental step size resulting from one power-stroke, and a 4 nm step representing a burst of four ATP hydrolysis events (Maillard et al., 2011; Abuin-Tam et al., 2011). The observation from Cordova et al. challenges this hypothesis. Consequently, optical tweezers data were not conclusive for the argument that ClpX is able to fire four subunits in a burst.

I would argue that the disagreement on the fundamental step size between the hand-over-hand model and the optical tweezers measurement could stem from a lack of understanding of the dynamics of ClpX. The evidence from optical tweezers should not be used to bluntly reject the hand-over-hand model. The hand-over-hand model does not provide enough information on such dynamics, but it offers a good framework for designing experiments for hypothesis testing.

Fei et al. (2020) proposed several alternatives to the hand-over-hand model in order to reconciliate the structural data with the observed ClpX step size and its capability to bypass ATPase mutant subunits. In their proposed models, pore loops from multiple ClpX subunits at the bottom of the spiral staircase move to the top of the staircase in a coordinated movement, thus allowing a much larger step size. Therefore, the models proposed by Fei et al. are similar to the hand-over-hand model, where only a single tyrosine loop could move during each duty cycle. However, a crucial variable that separates the models proposed by Fel et al. from the hand-over-hand model is the order of ATPase hydrolysis in the ClpX hexamer. In one of their models, stochastic hydrolysis of ATP in any subunit could build up strain in the hexameric ring, which is released as power-strokes. In another model, the conformational change is caused by ATP hydrolysis at the top of the spiral staircase (in the hand-over-hand model, hydrolysis happens at the bottom of the staircase).

The proposals from Fei et al. highlight the fact that the conformational change of ClpX and its ATPase cycle have not yet been studied in an integrated way. Importantly, Fel et al. could not consolidate their proposed hypotheses into a single model, reflecting the uncertainties of the integration between ClpX ATPase cycle and the conformational change. The models proposed by Fei et al. can be viewed as attempts to bridge the gap between the two aspects of ClpX. However, these proposed models have not been proposed for any other AAA+ ATPases, thus setting ClpX apart from similar AAA+ proteases like proteasomes. In their models, the coordinated movements of multiple tyrosine loops—which would require the disengagement of multiple tyrosine loops from

the substrate in order form them to move freely—could also lead to a significant reduction of grip between ClpX and the substrate during each duty cycle. Although these models address some problems of the hand-over-hand model, they at the same time raise new uncertainties that cannot be addressed yet. For this reason, in this thesis I adhere to the hand-over-hand model and acknowledge its problems.

Therefore, based on previously published ATPase mutant studies, I propose that during the burst phase, the consecutive subunits hydrolyze ATP in the clockwise direction, and the release of phosphate provides energy for each power-stroke. This model is an attempt to conform mutational studies with the hand-over-hand model concluded from structural studies, which readily rationalizes how consecutive subunits could hydrolyze ATP.

There are still unresolved problems that require investigation in future studies. The biggest problem is the conflict between a sequential firing order and the apparent functionality of several types of ATPase mutants. To solve this contradiction, there needs to be a mechanism that allows ClpX to bypass the defective subunits, but it is unclear how this is achieved.

The second problem of the model is that some predictions are based on the R370K mutation. R370K mutation is known to affect ClpX binding to ClpP. As ClpP association is mediated by IGF loops of individual ClpX subunits, when modifying these subunits with R370K mutation, it cannot be assumed that the overall dynamics between ClpX and ClpP remain unchanged in each case. Without knowing how R370K affects ClpP binding and dynamics, using this mutation runs the risk of confounding different effects.

The third problem is that this model only describes what happens during a single burst-dwell cycle, and it is not clear how ClpX resets itself after each cycle. The asymmetrical arrangement of the spiral staircase means that each ClpX subunit has a unique position in the spiral, and thus a unique role in the duty cycle.

In summary, after reinterpretation of the previous ATPase mutation experiments in light of the new ClpX structures, the evidence supports a clockwise sequential firing model with counterclockwise ATP loading, but it is unknown how ClpX bypasses a non-functional ATPase. The implicated clockwise ATPase-firing order agrees well with an F-to-A conformational change that has been proposed by the hand-over-hand model.

1.3.4 ClpP affects ClpX ATPase activity

It is well established that binding of ClpX to ClpP changes both how ClpX processes substrates as well as its ATPase activity. In Section 1.4, I will discuss changes of the unfoldase activity in detail. In this subsection, I briefly discuss the impact of ClpP on ClpX ATPase activity, and its potential influence on ClpX dynamics.

The IGF loops of ClpX bind to the empty hydrophobic clefts around the gate of ClpP (Section 1.2.4). In the absence of a protein substrate, gradually increasing ClpP concentration reduces the ATP consumption of ClpX when measured in bulk assays (Joshi et al., 2004). This result suggests that association with ClpP brings a change to the ClpX ATPase cycle. It is clear from structural data that the ATPase activity is regulated by multiple conserved motifs when ClpP is present, and some of these motifs might act differently in the absence of ClpP. However, there is no cryo-EM structure of ClpX without ClpP to inform this comparison, and there is the possibility that the crystal structure of ClpX might represent a low energy state that is not found under physiological conditions. Therefore, without structural data, the understanding of this effect is still lacking.

From the available ClpXP structure, the contacts between ClpX and ClpP are limited to IGF loops and potentially Pore-2 loops from ClpX. Both are very flexible in the cryo-EM structures, making allostery a less likely explanation.

On the topic of how ClpX ATPase cycle might regulate binding to ClpP, Alvaro Amor et al. in 2016 (Amor et al., 2016) discovered that in the absence of protein substrates, ClpX affinity for ClpP increases as ATP concentration rises (although ATP activity is reduced as a result of ClpP binding), indicating that it is ATP binding that leads to conformational changes that in turn prepares the IGF loops for binding to ClpP. This is further confirmed with a follow up study by Amor et al published in 2019, where they utilized a fluorescence quenching probe to measure the relative proximity of all IGF loops. They demonstrated that ATP binding brings IGF loop closer (Amor et al., 2019). On the other hand, depletion of ATP disengages ClpP quickly, implying that during an ATPase cycle ClpX is unlikely to hydrolyze all bound ATP, or otherwise ClpP might be released (Amor et al., 2016).

Amor et al. in the 2019 study also showed that the length IGF loop affects ClpX activity. Truncating IGF loop from its C-terminal end by three amino acids has no obvious effect on substrate degradation activity; whereas deleting two amino acids from the N-terminal side leads to significant defects in degradation activity. Adding flexibility to the loops by inserting three amino acid residues at the C-terminal side also leads to a degradation activity defect. For these ClpX mutants, Amor et al. did not report their affinities to ClpP,

nor the ATPase activities of ClpXP, making it difficult to judge the full implications of these modifications.

In summary, ATP binding of ClpX regulates ClpX affinity for ClpP, and ClpP is capable of influencing activities of ClpX via the bound IGF loops, but details of the latter process are still lacking.

1.4 Interaction with protein substrate

The most basic motion of ClpX is the motion of its pore loops paddling the substrate through its central pore. Because ClpX needs to process a wide range of substrates, the pore loops must be extraordinarily versatile. In section 1.2, I reviewed the ATPase firing cycle of ClpX, which governs the dynamics and kinetics of conformational changes. In this section, I discuss how ClpX interacts with substrates, and the gaps in the knowledge of the unfolding process.

In 1.4.1, I briefly introduce the background on the ssrA system which has been extensively used for targeting substrates to ClpXP in degradation assays. In 1.4.2, I review the three basic stages of substrate degradation that are mediated by ClpX—binding, unfolding, and translocation. In 1.4.3, I bring the focus to the interaction between the Pore 1 loops of ClpX and the substrate, and analyze the published studies on Pore 1 loop mutations. Here, I propose that it is advantageous to define grip, slippage and release in more specific terms, and use rate constants $\boldsymbol{k_{deg}}$ and $\boldsymbol{k_{rel}}$ to describe the degradation process in kinetic studies. In 1.4.4, I shift the focus to the amino acid composition of the substrate, and discuss how different substrate sequences may change $\boldsymbol{k_{deg}}$ and $\boldsymbol{k_{rel}}$. I argue that the exact mechanism by which certain sequences impair degradation is still waiting to be discovered.

1.4.1 The ssrA tag system for protein degradation

Unlike the ubiquitin-proteasome system, bacteria devolve the duty of the protein degradation into several proteasome-like particles, each assembled from a AAA+ ATPase module and a protease module. ClpP is a protease module that contributes a large portion of proteolysis in *E. coli*. Several AAA+ ATPases (including ClpX) are capable of binding to ClpP, and each AAA+ ATPase also has its own recognition signal system. The ssrA tag is one of the degradation recognition systems recognized by ClpX.

We use the ssrA tag as an available artificial means for delivering substrates to ClpX, without having a special interest in its biological role. Nonetheless, knowing the native function of the ssrA tag is helpful for interpreting results of biochemical assays.

The biological function of the ssrA tag system is to rescue ribosomes stalled at the 3' end of mRNA. This could happen when a ribosome does not encounter a stop codon during translation, or when reading a cluster of rare codons. The ssrA tRNA charged with alanine binds to the P site, allowing the stalled ribosome to continue translation with the message

encoded by the ssrA tRNA. The nascent peptide thus contains a C-terminal tag translated from the ssrA tRNA. This 11 a.a. C-terminal tag (with the sequence AANDENYALAA) is referred to as the ssrA tag, and it has been shown that mutating the last C-terminal alanine pair to aspartates disrupts targeting by ClpX (Keiler et al., 1996; Gottesman et al., 1996; Roche et al., 1999).

Julia Flynn et al. (2001) studied mutations of the ssrA tag, and showed the multi-purpose properties of the tag. The three C-terminal residues, LAA, are sufficient for binding to ClpX, while the rest of the tag is used for targeting to ClpA, or binding to the ClpX adapter protein SspB (Flynn et al., 2001). Andreas Martin et al. used a cross-linking method to show that the ssrA tag interacts with the RKH loop, the Pore 1 loops, and the Pore 2 loops, with the RKH loop functioning as the initial recognition signal (Martin et al., 2008b). This property allows the ssrA tag to be used with ClpX-ΔN complexes, which lack the N-terminal domain that binds to adapter proteins such as SspB (Hersch et al., 2004).

1.4.2 The binding, unfolding and translocation of the substrates

The fundamental interaction between ClpX and a protein substrate can be simplified as an interaction between a polypeptide chain and the pore loops of ClpX. If viewing the polypeptide chain as a track and ClpX as a motor protein, the most basic movement of ClpX is to walk along the track. Following the convention of published literature, I use ClpX as the reference point, and the process whereby the polypeptide is threaded by ClpX is referred to as translocation of the substrate. In this thesis, translocation is described by the hand-over-hand model.

When the polypeptide chain contains a folded domain, ClpX can pull the folded domain towards the unfoldase by translocating its adjacent unstructured region. Once the folded domain reaches a sufficiently narrow region of ClpX, translocation will be halted, because the folded domain cannot physically enter the central pore of ClpX. From this point, ClpX will repeatedly pull the unstructured region of the substrate, until the substrate unfolds, at which time effective translocation resumes. This period, lasting from the initial impeded encounter of the folded domain with ClpX until its eventual unfolding, is referred to as the pre-unfolding dwell. It should be noted that unfolding a native domain can take multiple smaller steps, due to the presence of unfolding intermediates that each needs to be unfolded in turn.

Four basic steps are required for ssrA-mediated degradation of a single-folded-domain protein by ClpX (Figure 1.4.1). The C-terminal ssrA tag first needs to bind to ClpX. Once

bound, ClpX needs to translocate the 11 a.a. ssrA tag and any contiguous additional unstructured regions, until it reaches a folded domain. ClpX then attempts to unfold the substrate during the pre-unfolding dwell stage. After unfolding, ClpX translocates the now unstructured region until it reaches the next folded domain or the end of the polypeptide. The translocated substrate is degraded inside the ClpP barrel.

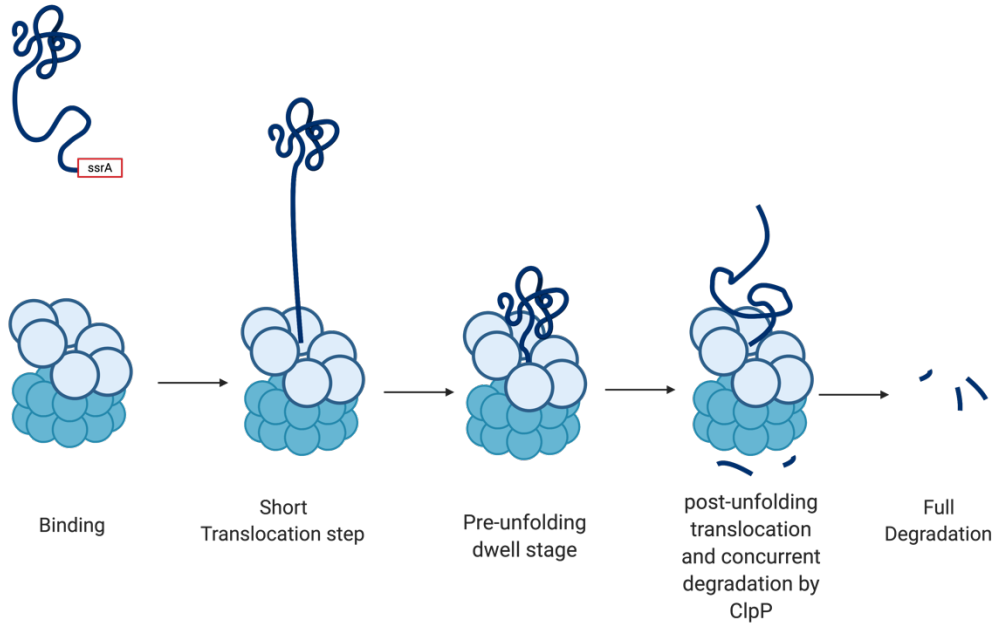


Figure 1.4.1 Stages of substrate degradation

Multiple optical tweezers studies have been conducted to characterize translocation and unfolding. The general observation is pre-unfolding dwell is the slowest step, with large variations even for the same type of substrate (Abuin-Tam et al., 2011; Cordova et al., 2014). The actual unfolding event when the substrate transitions from a folded state to a denatured state is very quick and is negligible in duration, likely reflecting the cooperativity of protein folding and unfolding dynamics. The average time of substrate unfolding is dependent on the stability of the substrate region adjacent to the unstructured area that ClpX grips, with longer unfolding time for more stable substrates (Cordova et al., 2014). To put the relative time frame in perspective, for substrates like GFP and the wild-type titin-I27 domain, the pre-unfolding dwell is typically measured in tens of seconds; the unfolding takes less than 50 milliseconds; the translocation time after unfolding is typically under 10 seconds (Sen et al., 2013; Cordova et al., 2014).

The ATPase activity of ClpX during pre-unfolding dwell stage cannot be easily measured in optical tweezers experiments. When using pre-denatured substrates and native substrates in bulk solution ATPase assays, Jon Kenniston et al. showed that ClpX hydrolyzes ATP faster when degrading pre-denatured substrates, indicating that the

presence of natively folded domain slows down ATPase activity **(**Kenniston et al., 2003). While it is still uncertain how ClpX ATPase adapts to the condition of the substrate, one appealing possibility is that the mechanochemical coupling property may lead to reduced ATPase activity when the movement of ClpX is stopped by an obstacle.

While single molecule studies like optical tweezers have extensively characterized the unfolding time and post-unfolding translocation time, Cordova et al. reported that the degradation rate measured in bulk solution is much slower (about 4 times slower in the case of GFP) than predicted from single molecule studies (Cordova et al., 2014). In the same study, Cordova et al. proposed a new parameter called "commitment", defined as the time calculated by subtracting the time for pre-unfolding dwell and post-unfolding translocation from the degradation time measured in bulk. When defined in this manner, commitment was shown to be the slowest step during substrate degradation. To put this into perspective, the pre-unfolding dwell for GFP lasts about 4.5 seconds but commitment takes 28 seconds on average.

It was shown from early bulk solution assays that when challenged by a substrate composed of multiple titin-I27 domains, ClpX might prematurely release partially processed substrate (Kenniston et al., 2005). The commitment step could therefore be explained as a result of high rate of premature release of the substrate before the eventual unfolding. In the formulation by Cordova et al., commitment is treated as a separate phenomenon from pre-unfolding dwell, but this might not necessarily be true. In the settings of optical tweezers, it is very difficult to quantify substrates that are released before unfolding, because a full dissociation of the substrate would terminate the measurement. \

Therefore, I think a kinetic model that takes into account the potential high off-rate constant can be used to as an alternative to the “commitment” model. In this model, one (or several) off-rate constant, k_{rel}, can be introduced to the kinetic scheme, as has been used by Kenniston et al. (2005), Koodithangal et al. (2009) and Kraut et al. (2012). This rate constant can be measured and compared between different substrates in combination with the forward rate constant, k_{deg}, without the requirement to characterize pre-unfolding dwell or post-unfolding translocation using optical tweezers. In this formulation, the net degradation can be characterized by the ratio between k_{deg} and k_{rel}.

While it is still an ongoing quest to understand factors that affect k_{deg} and k_{rel}, it is reasonable to predict that potential factors are related to both the inherent properties of ClpX (such as the ATPase cycle and the mechanical cycle), and extrinsic factors (such

as substrate stability, substrate amino acid sequence and so on). In the following subsections, I will cover the current understandings on the mechanism of substrate unfolding before describing a hypothesis.

1.4.3 The tyrosine loop and the unfolding of substrate

In order to investigate factors that affect k_{deg} and k_{rel}, it is important to look at both sides of the ClpX-substrate interaction during the unfolding process. In this section, I cover the effect of pore loops; in Section 1.4.4, I discuss the effects of substrate sequence.

While substrates have been shown to interact with RKH, Pore 1 and Pore 2 loops, the Pore 1 loop is the most likely site responsible for delivering the power-stroke of ClpX to the substrate. Among the four residues (GYVG) at the center of Pore 1 loop, both the tyrosine residue (Y153) and the adjacent valine residue (V154) have been shown to be important for substrate processing (Martin et al., 2008a). As mentioned in Section 1.2.3, both residues have been observed to form interactions with the substrate in the cryo-EM structures (Fei et al., 2020). While V154 primarily interacts with hydrophobic residues like alanine, the interactions of Y153 are harder to parse. Tyrosine is amphipathic, which allows its side chain to form hydrophobic, hydrogen-bonding or pi-stacking interactions (Koide and Sidhu, 2009). The bulky side chain might also drive the substrate forward with steric clashes. These properties certainly make evolutionary sense, because ClpXP are able to process a wide range of substrates with different sequences (Barkow et al., 2009). It is reasonable to expect that Pore 1 loop supports multiple modes of interaction with the substrate.

Andreas Martin et al. measured ATPase and degradation activity of the R-W-E/R-W-E pseudohexamer ClpX in combination with Y153A, V154F, or V154A mutations (R and E represents ATP hydrolysis mutants; R-W-E/R-W-E is assembled from two identical pseudo-trimers; details about the design of R-W-E/R-W-E construct can be found in Section 1.3.3) (Martin et al., 2008a). In general, the basal ATPase rate (measured without protein substrate nor ClpP) is changed in almost all cases, although to different degrees. This indicates that ATPase cycle can be modulated by dynamics at the Pore 1 loop. Specifically, Y153A and V154A mutations, which reduce crowding within the central pore, tend to cause elevated ATPase activity. Conversely, V154F mutation, with its larger sidechain, tends to reduce ATPase activity. These results suggest that ClpX ATPase has the inclination to run at higher rate when Pore-1 loops can move freely, and vice versa, when pore loops are restrained ATPase activity is slowed down accordingly. This prediction correlates well with the fact that when compared to processing pre-denatured

substrates, ClpX has lower ATPase activity when processing native substrates, since the folded domain will impede ClpX translocation movements before being unfolded (Kenniston et al., 2003).

Martin et al showed that among the tyrosine-loop mutants, ClpX with Y153A mutation requires a much higher amount of ATP to degrade a substrate, suggesting that Y153 is the key residue required to efficiently transfer the energy from ClpX power-strokes to the substrate. In a study published in 2015, Ohad Iosefson et al. conducted optical tweezers measurements on ClpX pseudohexamers with variable numbers of Y153A mutations (Iosefson et al., 2015). The experimental setup is similar to previous optical tweezers experiments described in Section 1.3.2. In the kinetic model proposed for substrate degradation, two off-rate constants were included to represent premature release of the substrates at different steps. Iosefson et al. pointed out that such release events would terminate the optical tweezers measurements, and therefore the off-rate was not reported from optical tweezers measurements. Here, adding more Y153A mutations (up to three out of six) in the hexamer increases the pre-unfolding dwell time for all substrates, but interestingly, these mutants have similar translocation speed and similar burst-dwell cycle (Iosefson et al., 2015). This result suggests forward degradation rate could be different, depending on whether the substrate is folded or not. Here, Y153A mutant would take longer to degrade a folded substrate but not a pre-denatured substrate.

It is not exactly clear from optical tweezers studies whether $\boldsymbol{k_{rel}}$ is affected by Y153A. Iosefson et al. reported two lines of evidence relevant to this question. First, slippages were observed in optical tweezers traces, where the substrate could be pulled out by the force applied by the optical tweezers. Compared to the wild-type ClpX, the mutant with two Y153A Pore-1 loops has an increased frequency of slippage during the pre-unfolding dwell, but the mutation does not cause higher slippage frequency during translocation. Second, Iosefson et al. used a non-degradable substrate—the methotrexate-bound *E.coli* DHFR-ssrA—as a competitive inhibitor for pre-denatured substrates, and measured the inhibitor constant, K_i, of this substrate on the Y153A mutants. They reported that K_i of DHFR-ssrA is low for wild-type ClpX, but is increased as more Y153A mutations are introduced. Their conclusion was that DHFR-ssrA binds better to wild-type ClpX, which agrees with the hypothesis that DHFR-ssrA is more likely to be released by Y153A mutants.

My own interpretation for observations on Y153A mutations is that the Y153A mutation reduces grip. A reduced grip won't impair degradation when little force is needed, as in the case of translocation of denatured proteins. However, when higher force is required, a reduction in grip would show a more prominent effect.

To elaborate this point further, grip can be defined as the static interaction between the tyrosine loops and the bound residues. Maintaining this interaction allows a pulling force from the tyrosine loop to be transferred to the substrate. A reduced grip, as in the case of the Y153A mutant, would make the interaction between Pore 1 loop and the substrate easier to break. More frequent interruptions of this interaction in turn reduces how much force ClpX can deliver to the substrate, thus impairing the unfolding of substrate. Because this interaction is prone to breakage, the model also explains why DHFR-ssrA substrate is more readily released from Y153A mutants.

In summary, current evidence points to the hypothesis that Y153A mutations cause slippage to be more frequent. For Y153A mutants, slippage likely happens when the weakened interaction between the Pore 1 loops and the substrate breaks under load.

1.4.4 The effect of substrate sequence on the interaction between ClpX and the substrate

Studies on Pore-1 loop mutations show that reduced grip of ClpX on the substrate could lead to premature release of substrate without degradation. For wild-type ClpX, an important question to ask is whether grip can be affected by the amino acid composition of the gripped substrate, and how k_{deg} and k_{rel} are affected. Many of the proposed hypotheses are inspired by protein substrates that exist in nature that can escape fully processive proteasome degradation. Such proteins have unique sequence features, such as glycine repeats, glycine-alanine repeats, or are rich in asparagine, serine and glutamine (Nassif et al., 2014). It is intuitive to speculate that the biophysical properties of glycine and alanine might disrupt the interaction of the substrate with the pore loop; these have, respectively, no side chain or a methyl, the smallest among native amino acids. It is less obvious how repeats of asparagine and such could prevent the overall degradation of the protein. In a study published in 2005, by analyzing the processing of transcription factors Cubitus interruptus (Ci) and NF-κB by proteasomes, Tian et al. (2005) proposed that low complexity sequence (defined by the SEG algorithm) adjacent to the folded domain impairs unfolding by unfoldases.

To understand the reason why low-complexity sequences may impair degradation, one essential step is to investigate whether different sequences of similar low-complexity have different or similar effects on k_{deg} and k_{rel}. On this subject, Daniel Kraut et al. in 2012 published a study on the effects of glycine-rich region from human p105 (GRR) and polyglutamine repeat (PolyQ) sequences on substrate degradation kinetics by proteasomes. The k_{deg} and k_{rel} for the degradation of a DHFR substrate with GRR

sequence was obtained by globally fitting the reaction curve with a kinetic model. However, the kinetic parameters for PolyQ substrates could not be determined by curve fitting. The conclusion from this study was that GRR sequence drastically reduces the forward rate constant k_{deg}, but has a very modest effect on k_{rel}; PolyQ similarly reduces the overall degradation rate, but the mechanism behind the effect is undetermined.

Priscilla Hiu-Mei Too et al. (2013) investigated how the glycine-alanine repeats (GAr), as found in the EBNA1 protein of Epstein-Barr viruses (Levltskaya et al., 1995), could disrupt ClpXP mediated substrates degradation (Too et al., 2013). This study was a continuation of previous work done in proteasomes (Hoyt et al., 2006). Both studies confirmed that the GAr works in tandem with a stable domain (in this case, the methotrexate-stabilized DHFR) to cause premature release of the substrate, and that an optimal spacing between the GAr and the folded domain can be found for both the proteasome and ClpXP. One interesting result from Too et al. was the off-rate of the substrate with GAr sequence, measured using a bulk solution chase experiment. Originally, Too et al. concluded that the GAr substrate had a higher rate of release from ClpX, but re-analysis of the same data by Kraut (Kraut, 2013) revealed a naive mistake in the original analysis. After reanalysis, the result showed that the presence of GAr does not significantly change the off-rate of the substrate, and the accumulation of the unprocessed intermediate was interpreted as a result of reduced forward rate constant, namely k_{deg}.

More recently, Tristan Bell et al. reported findings on interactions between ClpX pore loops and a wide range of substrate sequences during the unfolding step (Bell et al., 2019). In this study, Bell et al. discovered that among low complexity sequences, acid and basic sequences as well as GAr showed high levels of degradation just like high complexity sequences. Pure glycine sequence heavily disrupts degradation of the substrate, but adding a bulky tyrosine 3-5 a.a. from the folded domain restores the degradation to normal speed. The authors observed that ClpX might deliver its unfolding force mainly by interacting with the 4th residue away from the folded domain of the substrate. After replacing the 4th residue of the glycine test sequence with all other 19 amino acids, Bell et al. showed that ClpX favors amino acids with large, non-polar side chains at this position. They also demonstrated that while a single alanine replacement is not sufficient to promote degradation, replacing a second alanine anywhere between the 1st and 6th residue could significantly improve degradation.

In this laudable endeavor to test large combinations of low complexity sequences, Bell et al. have clearly demonstrated that the theory regarding low-complexity needs to be updated. It is possible that different low complexity sequences impair degradation by different mechanisms which cannot be generalized.

Their results also support the idea that the Pore-1 loops could support multiple modes of interaction with the substrate. The disruption brought about by the pure glycine test sequence can be rescued by either sprinkling the gripped area with multiple acidic or basic residues, or by introducing a single, strategically placed tyrosine or other large hydrophobic amino acid residue, or by two optimally spaced alanine residues.

It is not clear why a GAr sequence impedes degradation of the DHFR, but does not do so for the GFP substrate. It has been proposed by Hoyt et al. (2006) that the GAr sequence needs to collaborate with a mechanically stable substrate to inhibit its degradation. Hypothetically, it could be a combined function of how much force it is required to unfold the substrate, and how easy the interaction between the tyrosine loops and the bound area of the substrate can be broken.

Bell et al. does not show how a pure-glycine sequence inhibits degradation. Because the substrate with pure-glycine sequence is not sufficiently degraded *in vitro*, Bell et al. was not able to fit K_M and V_{max}. Therefore, there is no mechanistic insight on how this inhibition works. Without such insight, it is rather difficult to argue that the rescue of this glycine-only substrate by diverse amino acid replacements follows the same mechanism—namely by increasing grip, which is defined functionally by the authors as the ability to unfold substrates in this experimental setting.

In summary, it is still difficult to formulate a generalized hypothesis on the effects of amino acid composition and sequence on ClpX-substrate interactions. The results from Bell et al. demonstrate that several mechanisms might govern the process.

1.5 Current understandings, the unresolved questions, and a hypothesis for the release of substrate by ClpX

In the previous sections, I attempted to assess how well the hand-over-hand mechanism inferred from recent ClpXP structural studies fit previous biochemical and biophysical studies. I believe that the hand-over-hand mechanism can explain most of previous studies, with some notable exceptions.

The structural study showed that ClpX tyrosine loops adopt a spiral staircase, an arrangement shared by several AAA+ ATPases, indicating that a broadly conserved mechanism could be shared by these proteins. The nucleotide binding state from the cryo-EM structures also showed that the conformational change can be described by the hand-over-hand model, which is has also been proposed for other AAA+ ATPases.

The hand-over-hand model prescribes an orderly mechanism by which ClpX performs conformational change. However, ClpX has been shown to defy several predictions of this model. First, mutation studies on ClpX ATPase showed that ClpX is capable of bypassing the sequential order of conformational change. Second, step size of ClpX measured by optical trap system is larger than could be feasibly achieved by a strict hand-over-hand motion.

These highlight the gaps in the current knowledge of ClpX. One key question that structural analysis cannot address is the dynamics of ClpX. It could be argued that the deviation from the structural model suggests a dynamic mechanism not captured by structural studies. Importantly, there are several key areas of ClpX that show highly dynamic properties.

First, the pore loops are highly dynamic. These include the RHK loops, the Pore 1 loops with the conserved tyrosine, and the Pore 2 loops. In the case of ssrA mediated substrate degradation, the ssrA tag first engages the RHK loop, and is believed to be actively translocated into the central pore by Pore 1 loops.

Pore 1 loops have been observed in the cryo-EM structure to interact with substrate through both its conserved tyrosine residue and the adjacent valine residue. This suggests that tyrosine loops can grip different sequences by different modes of interactions. However, it is not exactly clear how these modes of interaction compare in affecting the forward degradation rate of the substrate, or its release rate.

Second, unlike the methodical translocase predicted by the hand-over-hand model, ClpX does not produce force in a uniformly orderly fashion, as shown in optical tweezer experiments. ClpX translocates substrates by cycles of burst-and-dwell. In a burst, ClpX rapidly moves the substrate for 1 to 4 nm distance, which is followed by a long dwell cycle. The exact size of the burst is not strictly coupled with ATPase activity, meaning that one ATP hydrolysis does not lead to 1 nm step, but more ATP hydrolyzed in a burst contributes to a larger burst.

Third, and most importantly, the dynamics of ClpX during the pre-unfolding dwell stage are poorly understood. The hand-over-hand model only describes the conformational change of ClpX for translocating the substrate, but it is not clear how conformational change of ClpX could be affected when translocation is impeded. It is clear that when ClpX reaches the folded domain after translocating an unstructured region, ClpX ATPase activity is reduced, but it is not clear whether this is because the burst size is reduced or the dwell between bursts is prolonged.

More importantly, while optical tweezers show that substrate is more readily released during pre-unfolding dwell than during translocation, the reason is not clear. Mechanical force is implicated in this phenomenon, as unfolding typically involves application of more force by ClpX than does translocation. However, what role force plays in causing substrate release is unknown.

I propose that further insights about ClpX can be gained by studying substrate release during unfolding. Factors that change how substrate could be released offer insight into the dynamics of the substrate unfolding process. Exploring such factors could reveal how force is transferred to the substrate, or the failure to do so. It could also reveal how processivity of substrate degradation is achieved, and how a loss of processivity could be mechanistically exploited for various biological processes.

Based on the findings from optical tweezers studies by Iosefson et al. (2015) and Maillard et al. (2011)., and the effects of tyrosine loop mutations on substrate degradation and release, I propose the following model that can be tested experimentally.

One key parameter that influences both substrate forward degradation rate and its release rate is the interaction between the tyrosine loops and the substrate. Specifically, increasing the strength of this interaction ensures that force delivered from ClpX is efficiently used for unfolding, while reducing its strength would cause the interaction to be easily broken.

When describing the processivity of substrate degradation in terms of k_{deg} and k_{rel}, which signify the degradation rate constant and the release rate constant of the same substrate, increasing the strength of interaction between tyrosine loops and the substrate should increase k_{deg} and reduce k_{rel} at the same time.

Findings from Too et al. (2013) and Kraut (2013) demonstrated that certain factors could reduce k_{deg} but leave k_{rel} unchanged. This indicates that the strength of interaction between tyrosine loops and the substrate should not be the only factor determining k_{deg} and k_{rel}. To fit these observations, the model can be changed by allowing that losing interaction between tyrosine loops and the substrate is only the first step that could lead to substrate release. In other words, loss of the interaction between tyrosine loops and the substrate does not necessarily lead to the release of substrate.

Therefore, the model predicts that there are factors affect substrate release rate without otherwise affecting the strength of interaction between the tyrosine loops and the gripped area in the substrate. Identifying factors that modulate k_{rel} would help gain new insights on the kinetic model of substrate unfolding.

Finally, the k_{rel} of substrate has not been extensively studied due to various technical challenges. This offers a unique opportunity to support the current understandings of ClpX from a different prospective. In this work, I explored the effects of substrate tail length, perturbations of the ATPase cycle, the role of ClpP, and substrate sequence on k_{rel}.

To study this problem, I have established a workflow for probing the substrate-ClpX complex life span using total internal reflection fluorescence microscopy (TIRF). In Chapter 2, the method for measuring the k_{rel} is documented. The reasoning behind the design, and the set of assumptions are covered in Chapter 3. Later in Chapter 3, I show that tail length can be used to modulate k_{rel} without perturbing the sequence of the substrate that the tyrosine loops interact. I also show evidence arguing that the unfolding force could destabilize the interaction between tyrosine loops and the substrate. Next, I show that the presence of ClpP could largely remove the effect of tail length on k_{rel}. Finally, I show that the direct measurement of k_{rel} for a substrate with GAr tail confirms the conclusions from Kraut et al. (2012) and Too et al. (2013) that GAr does not specifically change substrate release rate, but the surprising findings from recent cryo-EM structures prompts a discussion on how GAr could impede k_{deg}. In Chapter 4, I draw conclusions by comparing the results with previous studies and propose several hypothetical models. The work of this thesis answers some of previously unresolved questions, but raises a number of new interesting questions that will guide future studies on this subject.

Chapter 2
Materials and Methods

In this chapter, I cover the methods used for the experiment. The reasoning for the design and the validation of the methods are described in Sections 3.1-3.5

2.1 Purification of the proteins

The plasmids encoding biotinylated ClpX pseudohexamer (ClpX6B) and ClpP-6His was a gift from Dr. Philip Coffino (UCSF). ClpX6B was made by adding an Avi-tag at the 3' end of the ClpX6 linear hexamer from Martin et al. (2005). As noted in Section 1.3.3, ClpX6 was assembled from ClpX where the N-terminal ZBD domain was deleted (ClpX-ΔN). A monomeric ClpX-biotin was cloned by adding an Avi-tag at the 3' end of the ClpX-ΔN while removing the N-terminal his-tag to be used for the SPR experiment.

To study the substrate release by ClpX and ClpXP under TIRF microscopy, substrates were designed based on the same blueprint. Each contained an N-terminal his-tag, additional cysteine residues for conjugation of fluorophores, followed by a circular permutated GFP (cpGFP) domain which is easier to unfold than the native GFP, followed by the DHFR domain, followed by a variable test sequence, followed by the ssrA tag.

Specifically, starting at the N-terminus is an unstructured region containing both the 6his-tag and the reactive cysteines for conjugation of fluorophores. Cysteines were spaced at least 8 a.a. apart to reduce self-quenching by fluorophores; thus, the N-terminus sequence is MGCSSHHHHHHHSCSGLVPRGSCHMGGTS.

Following the his-tag is cpGFP domain, which is the same as the cp7-140-sfGFP variant described in Pédelacq et al. (2005). Degradation of this substrate has been characterized by Nager et al. (2011), where the equivalent variant was referred to as cp6-sfGFP-ssrA instead.

The cpGFP domain is followed by the *E. coli* variant of DHFR (*folA*). At the C-terminal end of the DHFR domain is the variable test sequence, where all variants listed in Table 2.1. The ssrA tag is placed at the C terminus of the substrate.

The test sequences were purchased as synthesized double-stranded DNA fragments, which were inserted between the DHFR domain and ssrA tag using NEBuilder HiFi DNA

Assembly kit (New England BioLabs) A list of the test sequences with the ssrA tag is shown in **Table 2.1**.

Table 2.1, tail sequences, including the test sequence and the ssrA tag, of substrates used in the experiment.

Substrate Name	Tail Sequence
11 a.a.-Tail	…**AANDENYALAA-COOH**
22 a.a.-Tail	…GLGARSAGITH**AANDENYALAA-COOH**
33 a.a.-Tail	…GLGARSAGITHLERPHRGLGD**AANDENYALAA-COOH**
48 a.a.-Tail	…GLGARSAGITHLERPHRGLGDISDQEAKPSTEDLGDK**AANDENYALAA-COOH**
GAr15	…GLGARSAGITHLERPHRGLGDIAGAGGGAGAGGAGGA**AANDENYALAA-COOH**
GAr37	…AGAGGAAGAGGGGAAGAAAAAGAGAGAGGAGAAAGGG**AANDENYALAA-COOH**
85 a.a-Tail	…GLGARSAGITHLERPHRGLGDISDQEAKPSTEDLGDK- GLGARSAGITHLERPHRGLGDISDQEAKPSTEDLGDK**AANDENYALAA-COOH**

(Test Sequence + ssrA tag; ssrA tag is **bolded**)

All proteins, including ClpXP and substrates, were expressed in the *E. coli* BLR(DE3), which is a *recA*$^{-}$ derivative strain of BL21 that may stabilize plasmids containing repetitive sequences (EMD Millipore). ClpX6B, which contains a C-terminal AviTag, were co-expressed with BirA (gift of Dr. P. Coffino) during expression in TB media in order to be biotinylated *in vivo*; all other proteins were expressed in LB media. Cells were grown at 37 °C until OD600 reached 0.4-0.5. For ClpP, cells were induced by 1 mM IPTG and harvested after 3 hours of 37 °C incubation. For all other proteins, cells were cooled gradually to 16 °C for 30 min. Expression of substrates was induced by 0.75 mM isopropyl β-D-1-thiogalactopyranoside (IPTG); 1 mM IPTG was used to induce ClpP, and 0.60 mM was used for inducing ClpX6B, the culture was supplemented with 50 μM D-biotin. Cultures were incubated overnight at 16 °C and harvested the next day.

For protein purifications, harvested cells were lysed using BugBuster Protein Extraction Reagent (EMD Millipore) with the manufacturer protocol. Briefly, BugBuster Reagent, a proprietary detergent-based lysis buffer, was supplemented with rLysozyme (EMD Millipore) at 5 KU per gram cell pellet and Benzonase (EMD Millipore) at 125U per gram cell pellet, 1 mM DTT and EDTA-free cOmplete Protease Cocktail (Roche). For purification of monomeric ClpX-biotin, 5 mM ATP and 10 mM MgCl2 were also added to the lysis buffer. The fully supplemented lysis buffer was then well mixed with harvested pellet at 5 mL per gram cell pellet. For ClpX6B and monomeric ClpX-biotin, lysis was carried out at 4-8 °C by incubating for 1 hour; For all other proteins, lysis was carried out

at room temperature for 20 min. The lysate was then clarified by centrifugation at 20000 rcf at 4 °C for 30 min.

Biotinylated ClpX was affinity purified using Pierce Monomeric Avidin Agarose (ThermoFisher) in a packed gravity flow column. Bound ClpX was washed with a wash buffer (100 mM PBS, pH 7, 150 mM NaCl), and eluted with the same buffer supplemented with 2 mM D-biotin. DHFR substrates and ClpP were purified using Ni-NTA Agarose (Qiagen) packed in gravity columns. Proteins were washed with a wash buffer (50 mM PBS, pH 8, 300 mM NaCl, 1 mM DTT) supplemented with 20 mM imidazole, and eluted with wash buffer supplemented with 250 mM imidazole. ClpX, ClpP were exchanged into HEPES storage buffer (25 mM HEPES, pH 7.4, 150 mM KCl, 10% glycerol) by first concentrating the elution fraction using Amicon centrifugation filter with repeated washes in HEPES storage buffer (EMD Millipore), which was then gel filtration using Sephadex G-25 in PD-10 Desalting Columns (GE Healthcare) and stored at -80 °C. DHFR substrates were exchanged into substrate storage buffer (25 mM HEPES, pH 7.0, 150 mM KCl, 10% glycerol, 1 mM DTT) using the same procedure.

Labeling of the DHFR substrates was done using the thiol-maleimide reaction. Sulfo-Cy3 and Sulfo-Cy5 were purchased from Lumiphore. Prior to labeling reaction, the 1mM DTT in the substrate storage buffer was removed using Zeba spin desalting columns (ThermoFisher). Protein to dye ration were kept at 1:10; reactions were conducted at room temperature for 2 hours in a vacuum desiccator. Reactions were quenched with excess of 2-mercaptoethanol 10x to the amount of dye molecules. The labeled proteins were buffer exchanged twice using Zeba spin desalting columns into HEPES storage buffer, concentrated with Amicon Ultra-0.5 centrifugation filter (EMD Millipore) and stored at -80 °C.

2.2 Biochemical assays

PD buffer containing 25 mM HEPES with pH at 7.4, 100 mM KCl, 20 mM MgCl2, 10% glycerol was used for all activity assays as well as TIRF imaging.

ATPase activity was measured using an NADH-coupled assay. The starting condition was with 1 mM NADH, 2.5 mM phosphoenolpyruvate, and 1/20 diluted pyruvate kinase/Lactic Dehydrogenase enzyme mix from Sigma-Aldrich (P0294), of which the final concentrations were 30-50 U/mL for PK, 45-70U/mL for LDH. Reactions were loaded in Corning UV-transparent half-area 96-well microplates and reduction of NADH was recorded using Tecan Spark 20 plate reader by measuring absorbance at 340 nm. The

pathlength was determined by first measuring the absorbance of NADH of known concentrations titrated from 1 mM to 100 μM at the same volume as the tested samples in the same type of microplate, and then fitting the Absorbance-Concentration curve by Abs340 = 6.22 (mM-1 cm-1) lightpath (cm) * [NADH] (mM). ATP hydrolysis rate was then measured by first determining the ΔAbs340/Δt for the sample, and then calculated by ΔAbs340/(6.22 * lightpath).

Fluorescence-based degradation assay was conducted in PD buffer. ATP was regenerated using 16 mM creatine phosphate and 3.6 U/mL creatine phosphokinase. Reactions were loaded in a Corning 96-well half-area black flat bottom polystyrene plate with non-binding-surface, and fluorescence were measured using Tecan Spark 20 plate reader by excitation at 485 nm and emission at 535 nm. A photobleaching control for each fluorescent substrate was included in the same measurement to correct for photobleaching caused by repeated readouts from the plate reader. Fluorescence was normalized by calculating the fraction between fluorescence at time-point t over fluorescence at time 0 **(F_t/F_0)**. In the bleaching control, the fraction of GFP that remains fluorescent at each time point was calculated using **($F_t^{Bleach\ Ctrl}/F_0^{Bleach\ Ctrl}$)**. The normalized fluorescence of the sample was then corrected for photobleaching by dividing the normalized level with **($F_t^{Bleach\ Ctrl}/F_0^{Bleach\ Ctrl}$)** ratio for each corresponding time point.

SPR was conducted using BioRad ProteOn SPR platform. The running buffer is based on the PD buffer described above, supplemented with 2 mM ATP, 10 mM MgCl2, and 0.05% BSA. Monomeric biotinylated ClpX was immobilized on a Neutravidin-decorated NLS chip, and a variant of the DHFR-substrate lacking the cpGFP domain and reactive cysteines was used.

2.3 Passivation of the coverslip, slides and the design of the flow cell

Flow cell entry and exit ports were drilled into the standard microscopy slides using a Dremel rotary tool with a 1/16 inch diamond drill bit. The layout was flexible according to experimental loads. Cutouts of channels were made on two layers of Parafilm, which was about 0.26 mm thick according to manufacturer specification. The entry and exit ports for each channel were 10 mm apart, and the channel was about 5 mm wide, 12 mm long. Thus, each channel could hold 10-15 μl volume.

Slides and coverslips were first washed with 1% Alconox solution in hot tap water. After thoroughly drying by air, all slides and coverslips were cleaned in a plasma cleaner with argon gas for 3 min. Immediately after plasma cleaning, the slides and coverslips were

incubated in base bath (5% KOH in isopropanol) for at least 2 hours at room temperature, but not longer than 4 hours. After this step, the wax coating of the coverslips from the manufacturing process should be completely removed. The slides were then rinsed several times using DI H2O, and was air dried. The cleaned class was then washed twice with hexane, and incubated with 0.3% dichlorodimethylsilane (DDS) dissolved in hexane for 2.5 hours at room temperature, protected from light. Before retrieving the glass, the slides and coverslips were briefly sonicated with the DDS solution to remove small insoluble particles attached on the glass. Excess DDS was then washed off with two more hexane washes post-reaction. Before storage, slides were gently rinsed in ultrapure water by repeated dipping. Cleaned slides were stored in 50 mL conical centrifuge tubes at -20 °C.

The flow cell was assembled before the imaging experiment. In the final product, the coverslip is facing down to be used with an inverted TIRF microscope; between the coverslip and the drilled slides is the Parafilm channel cutout. The port of the slides can be fitted with pipet tips.

To assemble the flow cell, the diameters of the ports were examined first. They had to be small enough to form a tight seal with friction-fitted plastic pipette tips without the tip protruding below the port into the flow cell; if the diameter was too large, the port was narrowed by inserting a pipet tip into the port and cutting off the exposed portion above and below the glass. The drilled ports on the microscopy slide were aligned with the Parafilm cutouts, and the coverslip was laid on top of the Parafilm cutouts, so that the Parafilm was gently pressed between the coverslip and the slide. The assembly was wrapped with aluminum foil and was heated on a 70 °C hot plate for 1 to 2 minutes. After the Parafilm melted, the assembly was gently pressed by hand to glue the glass together. The entire assembly was then cooled to RT. A buffer reservoir could be made by leaving a pipet tip at the entry port. Buffer was flowed in by withdrawing liquid from the exit port using a syringe.

2.4 TIRF imaging conditions

All steps in the TIRF experiments were performed in the PD buffer as used in biochemical experiments. First, 25 µl of 0.2 mg/mL biotinylated BSA was flowed into an empty channel and incubated for 5 minutes. Then, 0.2% Tween-20 was flowed in and incubated for 10 minutes. The channel was then washed with 50 µl of 0.01% Tween-20. 25 µl of 0.2 mg/mL Streptavidin in 0.01% Tween-20 was then flowed in the channel, and incubated for 1 minutes. Excess streptavidin was washed away by 50 µl of PD buffer. Then, the channel was equilibrated with 50 µl of ATP buffer or ATP/ClpP buffer, depending on experimental

requirements. 50 µl of 10 nM ClpX(P) pre-incubated with ATP (or ATPγS) was then flowed in the channel and incubated for 1 minutes, followed by a 50 µl wash with the same buffer used for equilibration. Finally, 20 µl of substrate sample mixture that were pre-heated to 30 °C flowed in, and the entry and exit ports of the flow cell were plugged.

TIRF microscopy was conducted using a modified Olympus IX-81 with an UApoN 100x/1.49 N.A. objective, 561-nm and 647-nm laser, and TRF89901 quad band set (Chroma) for Cy3 and Cy5 imaging. The critical angle was approximately 61.2° and the TIR angle was approximately between 64.65°and 68.60°, yielding a spatial constant for decay to 1/3 of the evanescent field between 124 and 88 nm. For imaging with Cy3, the power output of the 561-nm laser was set to 5 mW; for imaging with Cy5, the 647-nm laser was set to 10 mW. Images were captured with 100 ms shutter using a Hamamatsu Flash 4.0 camera. TIRF was achieved using an azimuthal TIR-FM method with a steerable mirror set up to minimize interference patterns (Johnson et al., 2012) The measurements were carried out at 30 °C.

2.5 TIRF data processing and analysis

The data processing was conducted using a custom-built Python script. Each frame of the time course movies was scanned for circular blobs using the Laplacian of Gaussian (LoG) method from a published library package (Van Der Walt et al., 2014). The edge sharpness of the blobs should be above a user-defined threshold to be collected by the function for subsequent processing; the value of this threshold was determined by gradually increasing the threshold until no more background camera noise was mistaken for a real blob in three randomly selected frames within the movie. In my experience, the numeric value of the blob-detection threshold has been consistent for all processed samples. The collected dots were then fitted with a 2D Gaussian function and assessed for its brightness and shape; dots of too high brightness, or not circular (defined by extreme ratios of its x and y widths) were filtered out at this step. The remaining dots were then tracked over frames to be compiled into traces for each punctum. The tracking method allowed a small number of gaps in each trace to account for problems like fluorophore-blinking or overly aggressive filtering in previous processing steps. Traces that start on the first frame or end on the last frame of the time course movie were filtered out, due to uncertainties of the lifetime of these puncta. Traces that start and end on the same frame were also filtered out, due to them having a higher likelihood of representing non-specific interactions or fluorescence impurities. For each imaging session, multiple 10-min movies were analyzed this way. The outputs, which contain the dwell time of each puncta, were pooled into a single database for each imaging condition.

Statistical tests and curve-fitting were done using MATLAB (R2019b distribution). The number of traces having the same dwell time were counted from the pooled trace databases of each imaging condition, and the tallies were sorted. The resulting distribution showing trace counts for each dwell time was fitted with a single exponential decay using $A*e^{(-t/\tau)}$, with the least-absolute-residual robustness mode turned on for measured traces in the absence of ClpP. The 95% confidence interval of the fitted coefficients were calculated using MATLAB's built-in function confint(). Histogram and curve-fiting plots were produced in MATLAB.

Chapter 3
Measuring Mean Lifetime of the ClpX-Substrate Complex with TIRF Microscopy

3.1 Introduction to the experimental design

To explore the factors that could affect k_{rel} of ClpX, we designed a TIRF microscopy workflow to directly measure the ClpX-Substrate interaction lifetime at single molecule level. The key feature of TIRF microscopy is that fluorophores need to be within a thin evanescent field, ~100 nm, above the coverslip in order to be excited efficiently. Fluorophores that are not attached to the cover slip are filtered out in part because of their greater distance, they are not excited as efficiently, and in part because they are moving, their fluorescent emission is spread spatially across many pixels. Thus, by immobilizing ClpX on the coverslip, the fluorophore-labeled substrates that bind to ClpX should be easily visualized by TIRF microscopy, while free substrates above the coverslip will be filtered out.

TIRF microscopy has been used successfully to study ClpXP kinetics in the past. Yongdae Shin et al. in 2009 published a study on ClpXP-substrate interaction using TIRF microscopy. In this study, the fluorophore-labeled substrate was engaged with ClpXP in a reaction mixture before loading the sample for microscopy measurements. During the sample loading process, the ClpXP-substrate complex was stabilized by EDTA and ATPγS. After immobilizing ClpXP-substrate complex on the microscopy slide via the biotin moiety on ClpX, EDTA and ATPγS were washed away by perfusing the flow cell with ATP buffer. In this experiment, degradation of the substrate was reflected by the disappearance of fluorescence from the labeled substrate. After a time-lapse measurement, the fraction of fluorescent puncta remaining in a FOV can be plotted and curve-fitted using a kinetic model. Notably, for this kinetic model, Shin et al. assumed that the release rate of the substrate would be the same during the three stages of the substrate degradation—pre-unfolding dwell, translocation after unfolding, and a termination step where the fluorophore is released from ClpP.

The main goal of the study by Shin et al. was to measure the pre-unfolding dwell time and degradation rate at single molecule level. The main conclusion from this study was that ClpX was able to processively degrade ssrA-tagged multi-domain proteins from C-terminus to N terminus, with more stable substrates taking longer to degrade.

Since 2009, several single molecule studies on ClpX have been published, and some assumptions from the above experiment should be modified to reflect findings from more recent studies. In their TIRF setup, the substrate contains multiple domains, which were used to study the processivity of the degradation. Rate of degradation was reflected by the loss of fluorescence puncta overtime, which reflected either the degradation of the full-length protein, the release of the protein, or a photobleaching event. The obtained distribution of puncta lifetime was globally fitted by a kinetic model where several rate constants were included. Specifically, they assumed that from the initial binding, to the unfolding of each domain, the probability of releasing the substrate was the same for all these steps.

As described in Section 1.4.4, substrate release, defined as a complete dissociation of substrate, could be triggered by breaking the interaction between the tyrosine loops and the substrate. In this hypothetical model, the combination of higher force delivered from ClpX and the impediment of substrate movement due to steric clashes (such as the case during the pre-unfolding dwell stage) would increase the likelihood of interrupting the interaction between tyrosine loops and the substrate. This hypothesis thus predicts that the unfolding step would have a much higher release rate than the translocation step. This hypothesis was supported by optical tweezers studies by Iosefson et al. (2015), where ClpX was shown to lose grip with the substrate at a higher frequency during the pre-unfolding dwell stage.

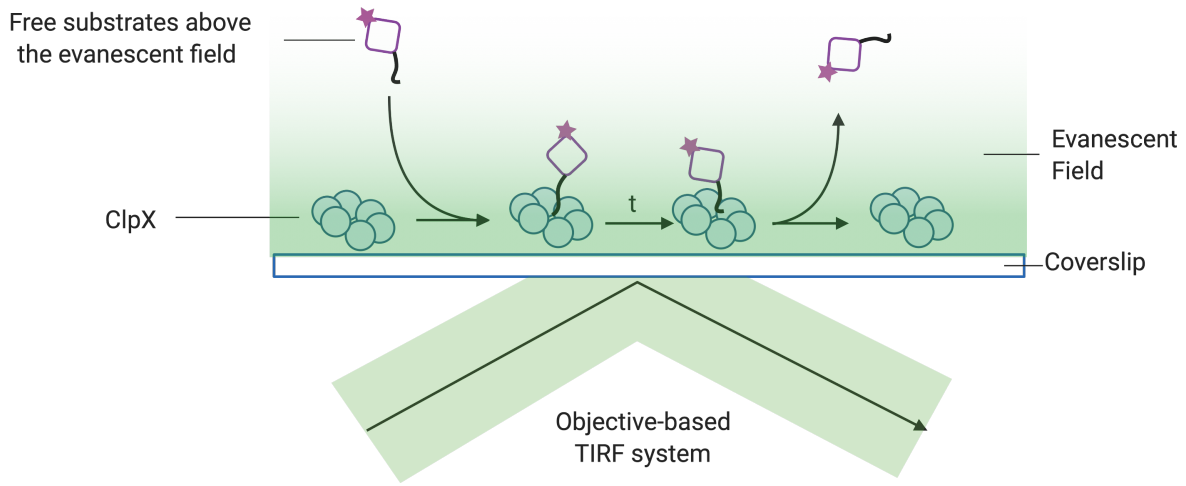


Figure 3.1.1 The TIRF assay measures lifetime of individual enzyme-substrate interactions

As a complement to optical tweezers studies, I propose to use TIRF microscopy to provide a different angle to tackle the problem. I plan to build on the work by Shin et al., and focus on measuring the substrate release rate using a conceptually simple TIRF microscopy setup. To overcome the potential bias introduced by pre-engaging the substrate, the ClpX-substrate lifetime can be measured in a pseudo-equilibrium environment. Here, a constant background of substrate is present in the solution; substrate could engage and disengage immobilized ClpX in an unsynchronized fashion, and lifetime is measured for each individual event (Figure 3.1.1).

To simplify the kinetic model, a substrate that cannot be unfolded by ClpX is used. To illustrate this point, a comparison between two hypothetical models is shown in Figure 3.1.2 and 3.1.3.

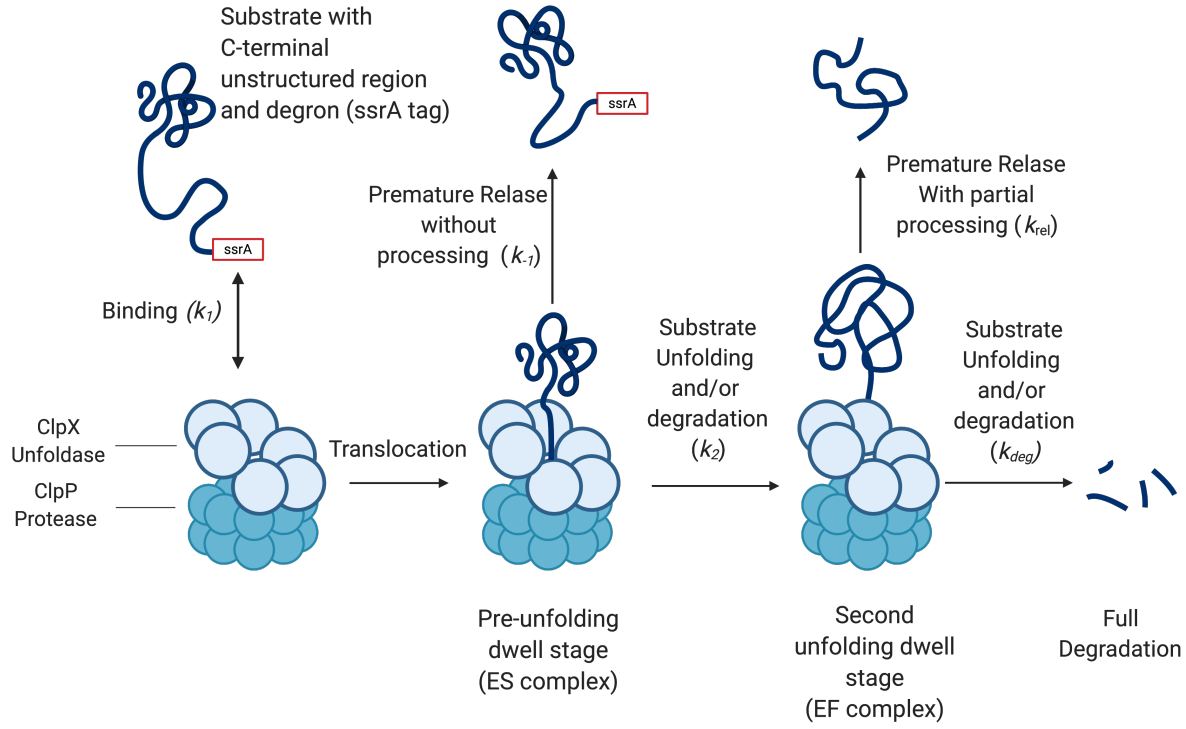
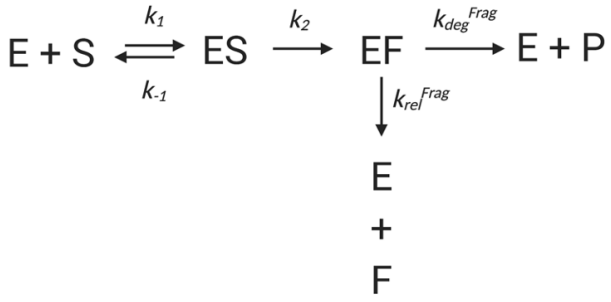


Figure 3.1.2
An example of model of single domain protein degradation kinetics

In the first model, shown in Figure 3.1.2, a substrate can undergo unfolding and degradation. Because my TIRF assay is designed to only measure the dwell time of puncta, the disappearance of a fluorescently-labeled substrate on the coverslip surface can have several possible causes, including the release of the substrate before unfolding, the release of a partially processed substrate, the release of a completely processed

substrate, and photobleaching. Here, “release” is defined as the dissociation of the substrate. In this example, it is assumed that a substrate has a single domain with a single intermediate during unfolding (one example is GFP, which have been demonstrated to have two steps in the unfolding process by ClpX according to Maillard et al., 2011). In this model,



without considering the effect of photobleaching, the speed of signal disappearance measured in enzyme-concentration per unit time, is thus dictated by the sum of the following: V = k_{-1}[ES] + k_{rel}^{Frag}[EF] + k_{del}^{Frag}[EF], where ES and EF denotes two stages of the enzyme-substrate complex. Because the ratio of [ES] and [EF] species is unknown, it has to be inferred by the complex function expressed using the total enzyme concentration, e_0, and total substrate concentration a_1. In this way, the expression becomes

$$V = e_0 \frac{a_1 k_1 (k_{-1} + k_2)\left(k_{rel}^{Frag} + k_{deg}^{Frag}\right)}{a_1 k_1 k_2 + a_1 k_1 \left(k_{rel}^{Frag} + k_{deg}^{Frag}\right) + k_{-1}\left(k_{rel}^{Frag} + k_{deg}^{Frag}\right) + k_2 \left(k_{rel}^{Frag} + k_{deg}^{Frag}\right)}$$

Apparently, in this mode the two processes dictated by k_{rel}^{Frag} and k_{del}^{Frag} are parallel reactions. It calls for methods like an elaborate design of the substrate to separately detect the presence of EF complex from the ES complex. At the same time, it also requires a method to distinguish the branching fate of the fragment bound on the EF complex—namely between the complete degradation from the release of the partially processed substrate. Without such methods, the best that can be achieved with TIRF microscopy measurement is the sum of k_{rel}^{Frag} and k_{del}^{Frag}. Therefore, I propose to use a simplified model where the substrate cannot undergo the unfolding process, as shown in Figure 3.1.3.

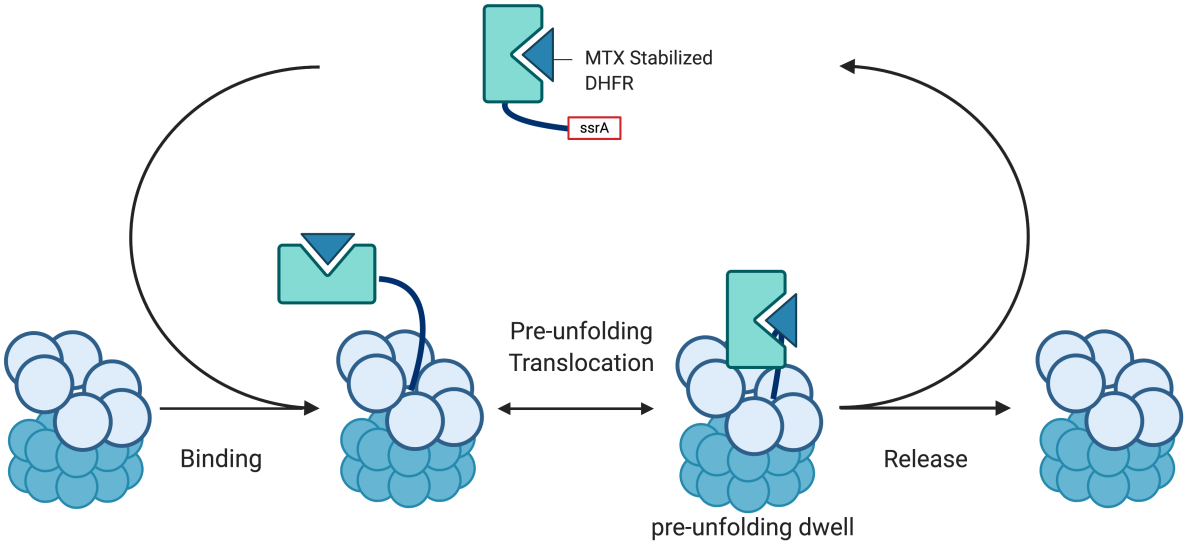


Figure 3.1.3
An example of a simplified kinetic scheme with non-degradable substrate

In an ideal system, the only kinetic pathway for the non-degradable substrate is binding and release. This means that the decrease of ES complexes is a first-order reaction, where,

$$E + S \underset{k_{rel}}{\overset{k_1}{\rightleftharpoons}} ES$$

and the rate can be written as:

$$V = \frac{d[ES]}{dt} = k_{rel}[ES]$$

One benefit of this model is that the identity of puncta seen under TIRF is unambiguous even when the substrate is labeled with a single color, because here all puncta represent the ES complexes. When not considering the effect of photobleaching under ideal conditions, if a punctum disappears, this signifies that the substrate is released from the enzyme and diffuses into the background. The lifetime of a punctum thus represents the lifetime of the ES complex measured at single molecule level.

To relate the bulk solution kinetics parameter k_{rel} to mean lifetime τ which describes single molecule events, calculations are presented, following a textbook example for single particle decay process.

For a first-order reaction, change of [ES] follows a single exponential decay function when measured in bulk solution:

$$[ES](t) = [ES]_0 \, e^{-k_{rel}t}$$

where $[ES]_0$ represents the starting concentration of [ES].

At single molecule level, the lifetimes of individual ES complexes are measured by TIRF microscopy, and the resulting distribution is fitted with an exponential model:

$$N(t) = e^{-\frac{t}{\tau}} N_0$$

where $N(t)$ marks the number of ES complexes that lasted time t before substrate is dissociated; N_0 represents the total number of ES complexes measured; τ represents the mean lifetime of the measured ES complex. The relation of τ and k_{rel} can be seen by normalizing the bulk solution model to a probability distribution, where the normalization factor, **c**, can be calculated as:

$$1 = \int_0^\infty c \, [ES](t)dt = \int_0^\infty c \, [ES]_0 e^{-k_{rel}t} dt = c \, \frac{[ES]_0}{k_{rel}}$$

Thus, the bulk solution degradation model can be normalized and expressed as a probability density function described by k_{rel}:

$$P(t) = c \, [ES]_0 e^{-k_{rel}t} = k_{rel} e^{-k_{rel}t}$$

the above probability distribution is an exponential probability distribution with random variable t and rate parameter k_{rel}, this shows that the original decay function is related to an exponential probability distribution by a scalar value **c**. Thus, the expected value for t, which is the mean lifetime, is given by the formula

$$\tau = \int_0^\infty t \cdot c \cdot [ES]_0 e^{-k_{rel}t} dt = 1/k_{rel}$$

where the mean time $\boldsymbol{\tau}$, is calculated by the formula that evaluates the mean of a distribution. In other words, the above calculations show that individual lifetime of the first order reaction is exponentially distributed, and the mean lifetime, $\boldsymbol{\tau}$, is given by the inverse of $\boldsymbol{k_{rel}}$.

Therefore, $\boldsymbol{k_{rel}}$ was obtained at single molecule level by fitting the distribution of ES complex lifetimes measured at single molecule level using the model below and taking the inverse of $\boldsymbol{\tau}$.

$$F(t) = Ae^{-t/\tau}$$

When adding the effect of photobleaching to the equation, for individual puncta it is not possible to determine whether the disappearance of a punctum is due to photobleaching or release of the fluorophore-labeled substrate. However, when plotting the distribution of dwell time, it is much easier to assess the effect of photobleaching at populational level. As will be discussed in the following section, by using techniques like increasing number of fluorophores per molecule to reducing oxygen level in the solution, I could not detect the effect of photobleaching on changing the distribution of ES lifetime (the result is shown in Section 3.5 Figure 3.5.5)

To conduct such a TIRF experiment, there are several technical issues to address. First, because the experiment requires a sufficiently large pool of free substrates present above the evanescent field, the passivation of the coverslip must resist nonspecific bindings of substrates at high concentrations. Second, the substrate needs to be non-degradable for TIRF microscopy for measuring its $\boldsymbol{k_{rel}}$, but it would be advantageous to be able to measure the forward degradation rate of the same substrate in a bulk solution assay. Third, the fluorescence signal should have sufficient longevity in a time-course study, in case it takes the substrate a long time to be released. Finally, an optimization of the puncta density is required to ensure that the field of view is not overly crowded in order to be accurately analyzed by an analysis program. These issues will be covered in the following subsections.

3.2 The passivation of coverslips

In 2014, Boyang Hua et al. published a novel passivation method. The glass surface is covalently coated with dichlorodimethylsilane (DDS), which renders the glass surface hydrophobic. This allows a layer of Tween-20 to self-assemble on the DDS-treated glass surface, which has a 30-fold greater passivation against proteins compared to a PEG-passivated surface (Hua et al., 2014). Because I intend to keep substrates freely diffusing on above the evanescent field to avoid pre-stabilizing the ClpX-substrate complex, having a higher passivation capacity is much more desirable.

I closely followed the protocol used by Hua et al. to prepare my coverslips and slides. I also followed their protocol for immobilizing biotinylated ClpX to the coverslip with only minor changes. Briefly, before adding Tween-20 on the coverslip, biotinylated-BSA is introduced first to bind to the surface. After formation of the Tween-20 layer, streptavidin is added to the surface to be immobilized by biotinylated-BSA. Biotinylated-ClpX is then added to bind to the bound streptavidin. For experiments requiring ClpP, biotinylated-ClpX and ClpP are pre-incubated with ATP before loading the sample; 100 nM of free ClpP is also present in the background during imaging to ensure that all ClpX particles remain charged with the protease particle. The details of this procedure, the design of the flow cell, and the TIRF microscope design can be found in Chapter 2. The immobilization strategy and the flow cell design are shown in Figure 3.2.1.

Because this was a very new passivation method, a series of tests were conducted to ensure that the protocol was valid. The first test was to challenge the passivation layer with high concentrations of proteins. In four flow cells, I immobilized four different levels of biotinylated-BSA on the coverslip surfaces, ranging from 0.2 mg/mL to total absence. I then challenged the passivated coverslip with 400 nM of Alexa488-labeled streptavidin in all four conditions. As shown in Figure 3.2.2, the puncta density reduces after decreasing immobilized biotinylated-BSA on the coverslip surface. The result suggests that 400 nM fluorescence-labeled streptavidin cannot non-specifically bind the coverslip, but can bind specifically to immobilized biotinylated BSA.

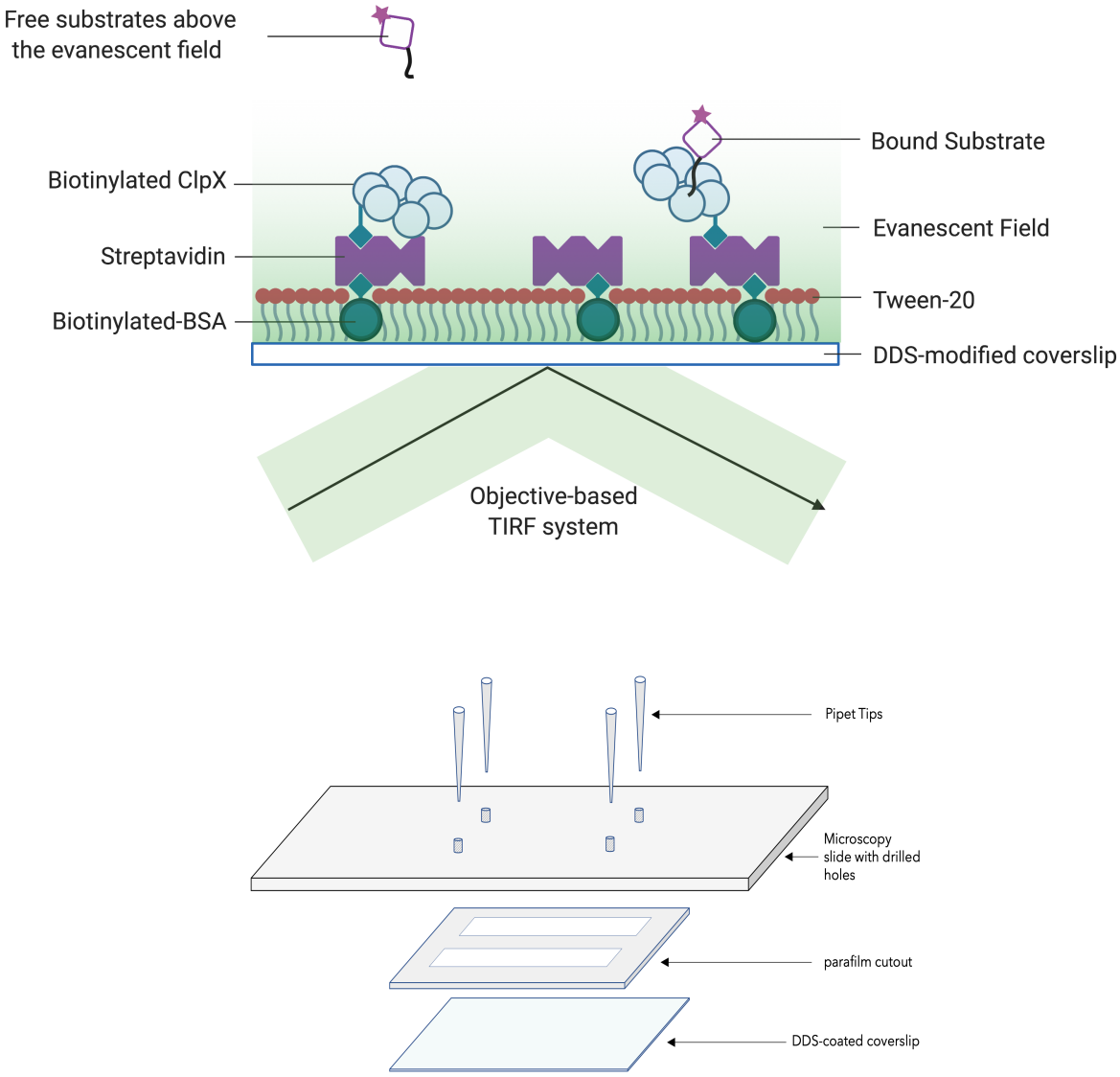


Figure 3.2.1

Top: Immobilization strategy for putting ClpX on the DDS-Tween-20 passivated coverslips to image ClpX-substrate complexes.

Bottom: the design of the flow cell used for TIRF microscopy. For a 15 mm x 5 mm flow cell cut on two layers of parafilm, the capacity of the channel is about 15 µl.

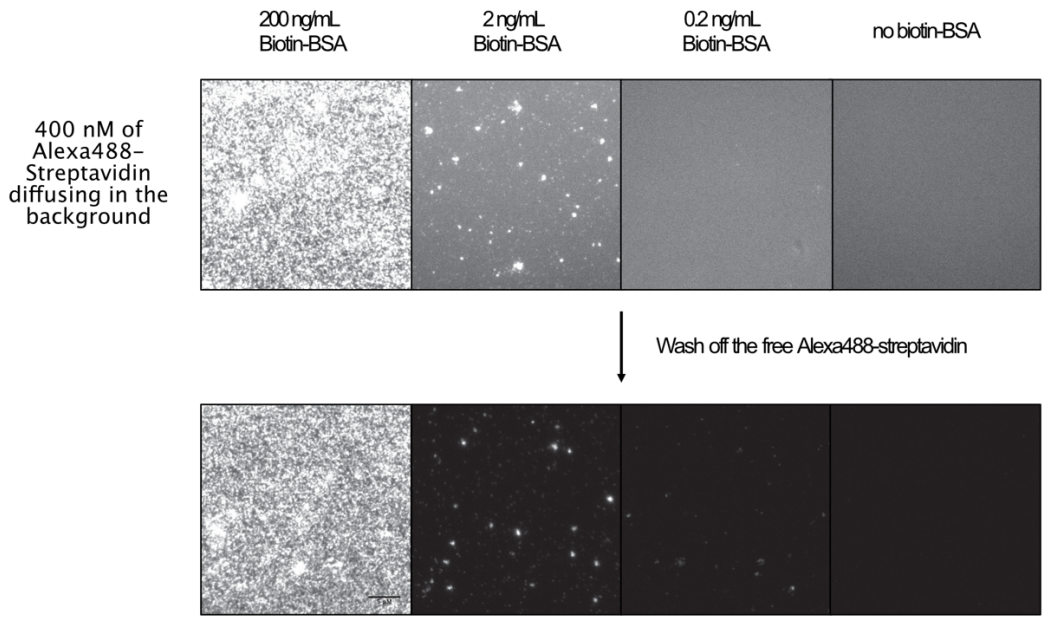


Figure 3.2.2 Challenging the passivated coverslip with 400 nM of Alexa488 Streptavidin. Fluorescent puncta from Alexa488 progressively reduced with the reduction of surface-immobilized biotinylated-BSA. A background glow can still be observed when the Alexa488-Streptavidin was allowed to diffuse in the background. After washing off the un-bound Alexa488-streptavidin, the background fluorescence was reduced, thus improving the contrast of the images. Total fluorescence at 200 ng/mL immobilized biotinylated-BSA condition was reduced to close to 0 when no biotinylated BSA was immobilized.

To test whether ClpXP was passivated equally well, I challenged the streptavidin-treated coverslip with 50 nM biotinylated-cy5-ClpX (with 100 nM ClpP in the background). When blocking the surface with excess free-biotin, the total number of puncta observed on a cover slip was reduced from 987 to 5. (Figure 3.2.3).

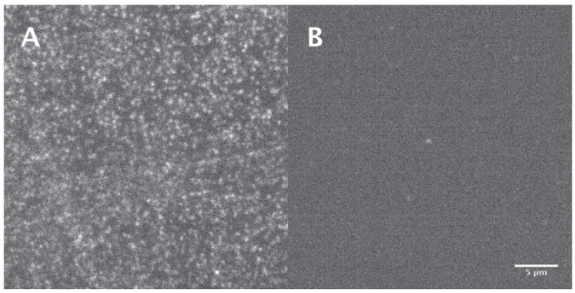


Figure 3.2.3. Comparison of binding of Cy5 labeled biotinylated ClpX in the absence (A) or presence (B) of excess free biotin. Immobilization was done by using 0.2 mg/mL BSA mix, 0.01 mg/mL streptavidin (ca. 400 nM) and 50 nM Cy5-ClpX. Puncta count is reduced from 987 to 5.

Next, to test whether ClpX could maintain activity under such conditions, ssrA-tagged substrates were introduced to test ClpX-specific binding. As introduced in Section 1.4.1, the three amino acids at the C-terminal end of the ssrA tag, LAA, are responsible for binding to ClpX. It was well established that mutating the last two amino acids from AA to DD abolishes ssrA mediated binding to ClpX (Keiler et al., 1996). In this thesis, a substrate of which the ssrA tag is mutated to DD is referred to as a ssrADD mutant. As a control for specificity in the passivation test, ssrADD mutation was used as a negative control.

In this specificity test, coverslip was challenged with a mixture of Alexa488-labeled ssrA substrate and Alexa594-labeled ssrADD mutant. As shown in Figure 3.2.4, compared to the number of puncta observed for the ssrADD substrate (30 were counted), the immobilized ClpXP was only able to capture the ssrA tagged substrate (1140 puncta were counted).

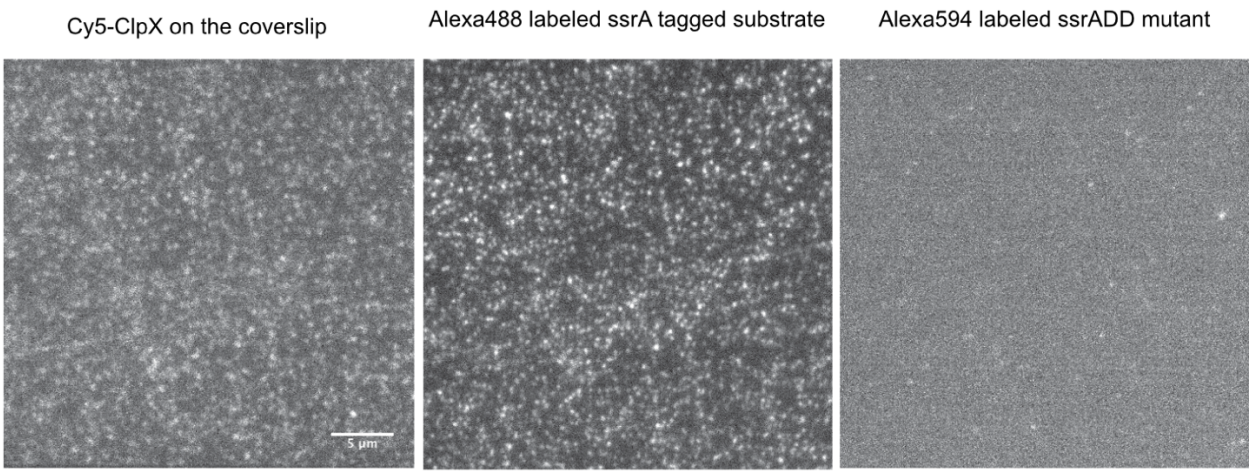


Figure 3.2.4. Comparison of binding for the ssrA-tagged substrate and the ssrADD mutant. ClpX were immobilized by 0.2 mg/mL biotinylated-BSA, 0.01 mg/mL streptavidin and 50 nM cy5-ClpX. 50 nM substrates were introduced. Compared to the ssrA-tagged substrate where 1140 puncta were counted, the ssrADD mutant channel had 30 puncta detected.

To assess how much of the result shown above was caused by the inferior optical properties of Alexa594, the labels between the ssrA substrate and the ssrADD mutant were swapped. In Figure 3.2.5, non-specific binding from the ssrADD mutant was minimal (4 puncta were counted) compared to the ssrA tagged substrate (475) even after switching the fluorophore between the substrates.

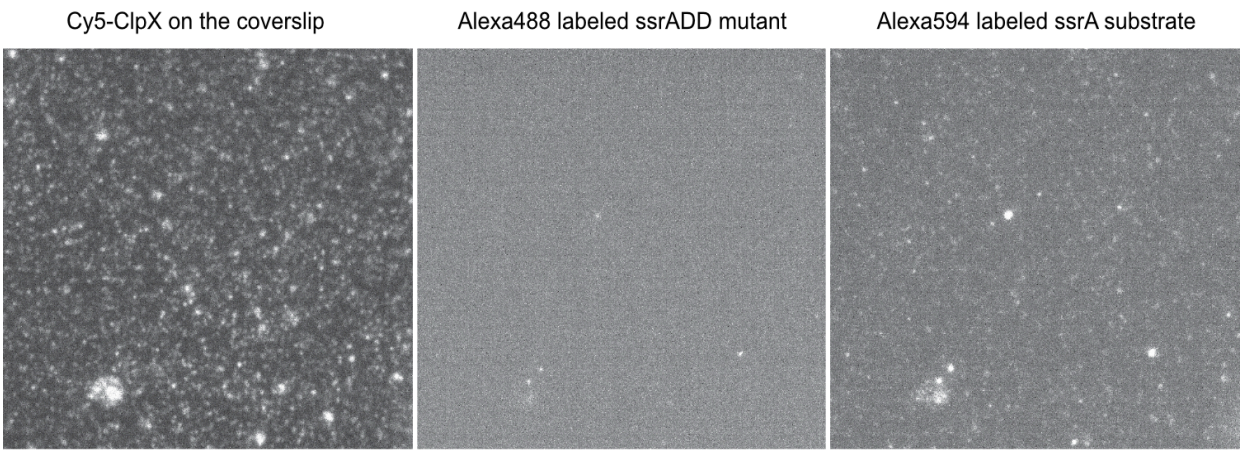


Figure 3.2.5. Comparison of binding for the ssrA-tagged substrate and the ssrADD mutant. Unlike in Figure 2.2.3, fluorophore was swapped between the ssrADD mutant and the ssrA substrate. ClpX were immobilized by 0.2 mg/mL biotinylated-BSA, 0.01 mg/mL streptavidin and 50 nM cy5-ClpX. 50 nM substrates were introduced. For the ssrA substrate, 475 puncta were counted; for ssrADD substrate, 4 puncta were counted.

It was clear that the puncta density was too high for proper single molecule studies. Puncta density could be reduced either by reducing ClpXP density on the coverslip or by reducing substrate in the background. We preferred the latter strategy for optical imaging. The former strategy would require a much higher substrate concentration in the background to ensure a sufficient number of observable binding events. This would significantly reduce the contrast of the obtained image even in the most extreme TIR angle.

In the second strategy, while the chance of binding to a specific ClpXP particle was low, having a large amount of ClpXP on the coverslip could ensure that at any given time enough binding events could be observed. This strategy also reduces the chance that the same puncta could be caused by multiple binding events, which could be useful for filtering purposes during data processing.

3.3 The design of the substrate

In the studies by Shin et al., a complex kinetic model was used to describe the measured TIRF data. It is understandable that such complexity was needed—multiple steps are required during substrate degradation, but it is impossible to distinguish these stages under TIRF microscopy when all of them manifest as puncta with identical characteristics.

Because our goal was to study substrate release rate (defined as the rate of dissociation of the substrate from ClpX), we decided to avoid substrate unfolding altogether. This strategy allowed us to model the interaction as a simple binding-and-release model with two kinetic parameters—the on-rate and off-rate. This was a departure from previous works on substrate release rates done by Kraut et al. (2012) and from Too et al. (2013), where partial degradation was required to accurately measure the substrate release in bulk solution experiments, but as discussed in Section 3.1, single molecule TIRF microscopy could directly record the lifetime of single enzyme-substrate complex, which was not easily measured in bulk solution assays. Therefore, the requirement of a substrate for single molecule assays would be different from that of the bulk assays. Nonetheless, we did see the advantage of being able to use the same substrate for both single-molecule work as well as bulk solution studies. The final design of the substrates included a fluorescent protein domain that could be used to track its degradation in bulk solution assays.

The end result of the above considerations was a series of substrates consisting of a modified circularly-permuted GFP (cpGFP) domain and an *E. coli* dihydrofolate reductase (DHFR) domain. This substrate is referred to as the "DHFR substrate" in this thesis, and the basic topology is shown in Figure 3.3.1. The ssrA tag is at the C-terminus of the DHFR substrate, and the substrate is purified by its N-terminal his-tag. The presumed direction of degradation by ClpXP is from the C-terminus to the N terminus. By itself, the *E. coli* DHFR domain is already highly resistant to ClpXP degradation. When bound to the drug methotrexate, the DHFR domain is further stabilized to be fully resistant to ClpXP degradation. In all the TIRF experiments I conducted, the DHFR domain was stabilized by MTX. In degradation assays measured in bulk, MTX was typically not included unless otherwise stated.

Between the DHFR domain and the ssrA tag, an unstructured amino acid sequence, referred to as the test sequence, could be inserted. The C-terminal unstructured region thus includes the test sequence and the ssrA tag. To simplify nomenclature, the "C-terminal unstructured region" is referred to as "tail region" or simply "tail" of the substrate in this thesis. When the non-degradable substrate binds to the ClpX central pore, the tail region of the substrate should be where ClpX pore loops primarily interact with. By switching test sequences, we can impose different types of amino acid features to interact with ClpX (Figure 3.3.2).

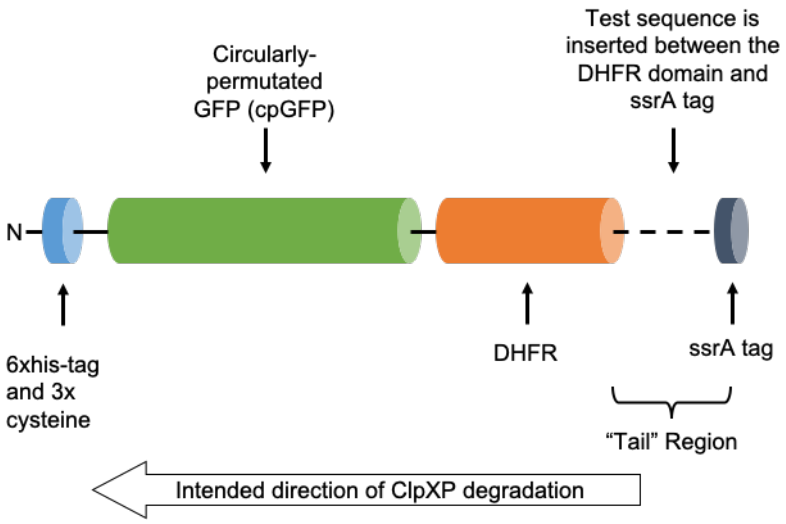


Figure 3.3.1.
The topology of the DHFR substrates designed for the TIRF experiments.

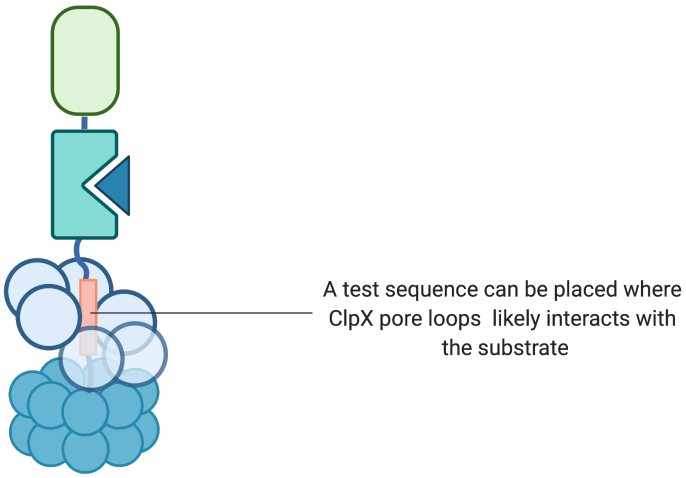


Figure 3.3.2
The placement of the test sequence is between a tightly folded domain the C-terminal degradation initiation signal, the ssrA tag, which ensures that the test sequence can be imposed on the central pore of ClpX

After binding the ssrA tag, ClpX in theory should rapidly translocate the substrate tail into the central pore and stop at the DHFR domain. Because the MTX-stabilized domain would prevent ClpX from proceeding further, ClpX should mainly interact with the tail region before releasing the substrate.

On the N-terminal side of the DHFR domain is the modified cpGFP domain. The cpGFP has been reported to be a very easy to unfold substrate (Nager et al., 2011). An HRV 3C protease cleavage site was inserted between the two C-terminal beta strands of the cpGFP domain to further destabilize the substrate, although the site was not utilized for this work. Therefore, this domain is only referred to as cpGFP in this thesis, but this feature could be put to use in future studies.

It was well established that ClpXP could degrade substrates from the N terminus, C terminus or an interior position, and could translocate substrate efficiently in both directions (Olivares et al., 2017). By having a low stability reporter domain on the N-terminal side of the DHFR domain, degradation initiated at N terminus can be detected by SDS-PAGE fractionation. If this happens, it would be problematic for the experimental design, because ClpX would bypass the test sequence inserted between the ssrA tag. Another use of the cpGFP domain is that when the DHFR domain is not stabilized by MTX, the GFP fluorescence could be used in a bulk assay to track the degradation of the full-length substrate. Although convenient, as will be discussed in detail at Section 3.8, the DHFR substrate was not very sensitive for detecting changes in degradation rate, due to its very low rate of degradation under even the most optimal condition.

In order to improve the brightness of the puncta at single molecule TIRF condition, the substrate was thiol-labeled with multiple copies of fluorophores by the three cysteine residues inserted in the N-terminal unstructured region. to facilitate labeling by thiol-reactive fluorophores. The cysteine residues were spaced by at least 8 a.a. apart to reduce self-quenching between fluorophores.

The substrate was purified using a his-tag at the amino-terminus. A carboxyl-terminal his-tag was not used because the ssrA tag must be positioned at the C-terminus to function as a degron. If a his-tag is inserted between the ssrA tag and the DHFR domain, this purification tag would act like a low-complexity test sequence (“low-complexity sequence” is introduced in Section 1.4) and interfere with the release rate measurements.

One undesirable outcome of this substrate design was that it was not easily purified in *E. coli*, especially when the test sequence could lead to increased partial degradation by ClpXP. Because the substrate was expressed with its ssrA tag, the native proteases in *E. coli* could still attempt to degrade the substrate. Without MTX stabilization, a small portion of the tail could be snipped off. When purified with the N-terminal his-tag, an impurity of DHFR substrates lacking ssrA tags would be co-purified. In one example shown in Figure 3.3.3, two substrates of different tail sequences were purified. One of two contained a small stretch of glycine alanine repeats at its tail, and was co-purified with two lower bands

containing the GFP domain. Nonetheless, these fragments purified by the amino-terminal his tag, did not contain the carboxyl terminal ssrA tag and thus did not target to ClpX nor ClpXP and were not imaged in TIRF and not further degraded by ClpXP. The ssrA-tagged species was quantified by an SDS-PAGE titration against a BSA standard.

The same degradation assay shown in Figure 3.3.3 D to G also showed that stabilization of the DHFR domain by MTX blocked the degradation of full-length protein. This supported that the directionality of the degradation of the DHFR substrate is from the carboxyl-terminal ssrA tag to the N terminus. This also confirmed that the MTX-stabilized DHFR domain was resistant to ClpXP mediated degradation (Lee et al., 2001).

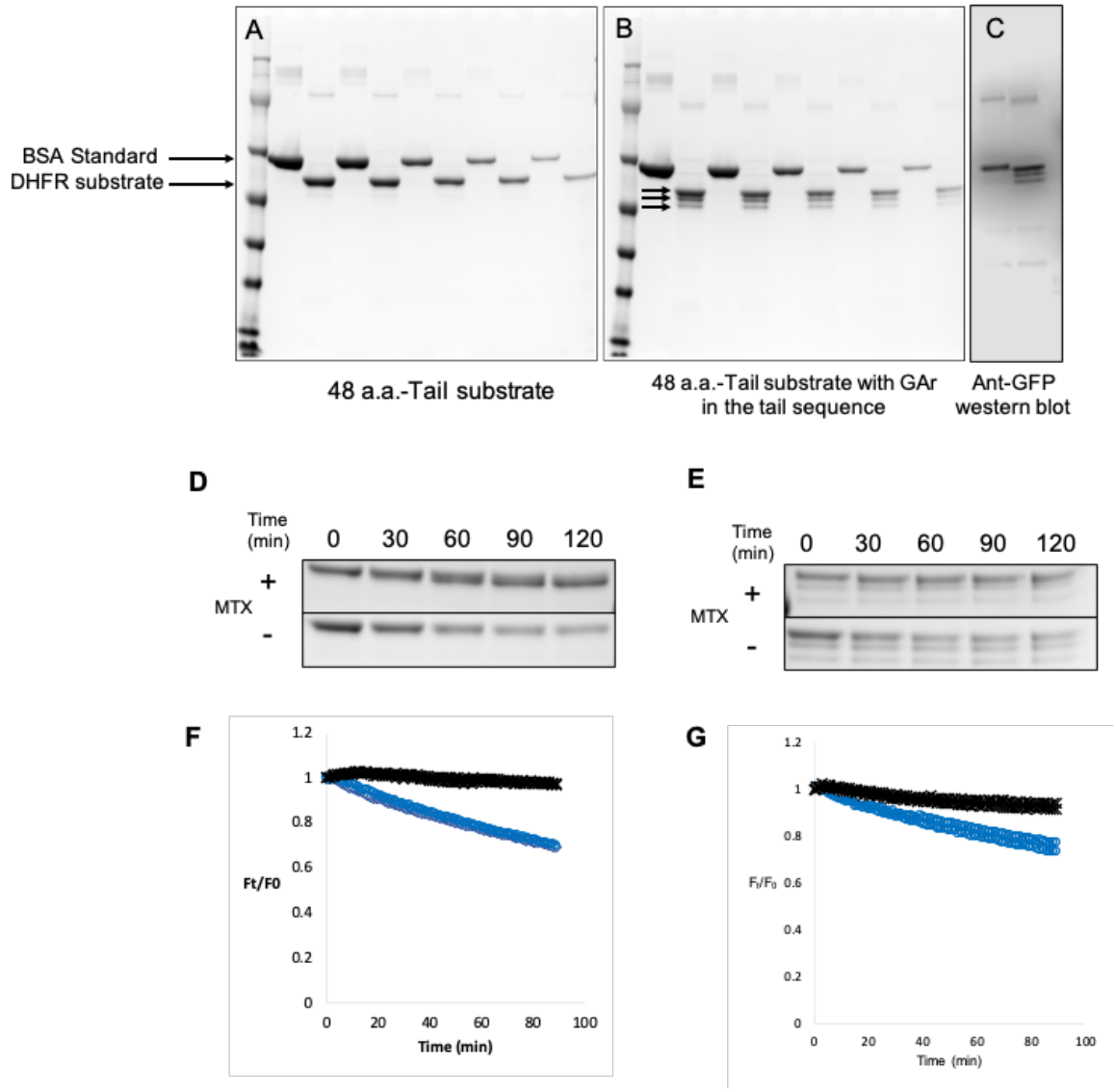


Figure 3.3.3

A-B: Coomassie staining of SDS-PAGE gels comparing of the band patterns of purified DHFR substrates with A) 48 a.a. tail, and B) 48 a.a.-tail containing glycine alanine repeats (GAr15; defined in tail sequence table in Chapter 2). In B), in addition to the full-length GAr substrate (indicated by the top arrow pointing at the main band at the top), two lower bands were co-purified (indicated by the other two arrows below the top band).

C: Anti-GFP Western blot of the same two substrates shown in A and B, which probes the cpGFP domain of the substrates. The blot shows that the lower bands for the GAr substrate are the same substrate, with their the C-terminal portion truncated.

D-E: Degradation of the 48 a.a.-tail substrates (D) and 48 a.a.- tail with GAr (E) by ClpXP in the presence or absence of 100 μM MTX. Gels were Coomassie stained. The presence of MTX effectively blocked degradation of the full-length product.

F-G: The same degradation assay shown in D and E, repeated by tracking the loss of the fluorescence from the cpGFP domain, which reflects the degradation of the full-length product. of the 48 a.a.-tail substrates (F) and 48 a.a.- tail with GAr (G) by ClpXP. In the presence of MTX (the series colored by black) degradation was blocked; in the absence of MTX (the series colored by blue), substrate was degraded slowly over time.

3.4 Imaging conditions and data processing routines

In the finalized experimental setup for TIRF imaging, I used high levels of biotinylated-BSA and streptavidin on the coverslip and 5 nM protein substrate in the background. The density of ClpX on the coverslip under this condition would be similar to that shown in Figure 3.2.2(A**)**, and only a small fraction of ClpX puncta would be bound with protein substrates due to the below-K_M condition. Fluorescently labeling ClpX under this condition in a multi-color TIRF experiment did not add advantages to data interpretation. Eventually, we settled on having only two fluorescent species in the TIRF imaging experiment—the DHFR-ssrA substrate labeled by Cy3, and a ssrADD mutant labeled by Cy5. The ssrADD mutant, which did not measurably bind to ClpXP, was used to map passivation imperfections on the coverslip, which should be avoided during imaging. We chose Cy3 for its great optical properties and longevity in the presence of an oxygen scavenging system (using the recipe from Aitken et al., 2008) and the antioxidant Trolox.

In a typical TIRF experiment, a flow cell was assembled by gluing the coverslip and slides using 70°C-melted Parafilm with channel cutouts. After the flow cell was treated with biotinylated BSA, Tween-20 and streptavidin, ClpX(P) was flowed into the flow cell, followed by a washing step; afterwards, a mixture containing MTX, the ssrA substrate, the ssrADD mutants, Trolox, and the oxygen scavenging system was flowed into the flow cell. The entry and exit ports of the flow cell were then plugged to reduce evaporation.

Imaging was conducted at 30 °C, with time courses taken at 2 sec intervals for 10 minutes. Each field of view was examined by the density of ssrADD puncta before taking the time course to ensure that passivation was functional. Typically, for each condition, about 5 such time courses were taken. Details of the buffer compositions, microscopy setup and data processing can be found in Chapter 2.

The dwell times of individual puncta were extracted from the time courses using a processing routine described in Chapter 2. The dwell time of the puncta was interpreted as the lifetime of individual ClpX-substrate complexes.

As discussed in Section 3.1, the distribution of the substrate dwell time is assumed to be governed by a single exponential function. After obtaining individual dwell times for ES complexes, ES complexes with the same lifetime are counted. The distribution of the count, sorted from 0 sec to 10 minutes, is fitted using the model A*e^(-t/τ), which gives the mean lifetime τ. The release rate $\boldsymbol{k_{rel}}$ is calculated as 1/τ**.**

3.5 Substrate release rate can be influenced by substrate tail regions outside of the gripped area

Based on findings from Bell et al. (2019), ClpX exerts unfolding force most efficiently on the amino acids at 3 to 5 a.a. away from the folded domain (discussed in Section 1.4.4). If substrate release were entirely determined by loss of static interaction between the tyrosine loops and the gripped area, extending the tail region should not affect the release rate at all. However, as I have proposed at the end of Chapter 1, loss of static interaction and the release of the substrate could be two separate events with some overlapping mechanisms. According to the results from Too et al. (2013), unprocessed substrate is released by backing out from the ClpX central pore, as covalently closing the hexameric ring with disulfide bonds does not kinetically affect the formation of degradation intermediates. One way to test my hypothesis is to measure the release rate of substrates from ClpX with different tail length. Based on my stated hypothesis, the expected result is that extending the tail length of the substrate should increase the opportunities for ClpX tyrosine loops to re-establish static interaction, thus reducing the rate of release, thereby extending the mean lifetime of the ClpX-substrate complex.

The substrates were designed so that spacers of different lengths can be added between the ssrA tag and the folded DHFR domain. For this work, substrates with tail lengths of 11 a.a., 22 a.a., 33 a.a., 48 a.a., or 85 a.a. (when length of ssrA tag is included) were made. The 85 a.a.-tail was created by doubling the spacer of the 48 a.a.-tail. Except for the 11 a.a.- tail substrate which only contains the ssrA tag, all other substrates share the same first 11 a.a. residues adjacent to the folded domain. Based on the published cryo-EM structures (Ripstein et al., 2019 and Fei et al., 2020) and the degradation studies (Bell et al., 2019), this 11 a.a. region is the likely location gripped by ClpX when it stalls at the folded domain.

Special consideration should be given for the 11 a.a.-tail, the sequence gripped by ClpX differs from other four substrates. As introduced in Chapter 1.4, the sequence complexity hypothesis, while not offering a mechanistic explanation, is a relatively good way to predict how degradation could be impaired. Bell et al. further showed that only glycine-rich sequences significantly inhibit degradation of GFP. The ssrA tag (AANDENYALAA) does not contain any glycine residues, and five bulky amino acid residues are placed in its 3-7th position, which, according to Bell et al. (2019)., should make ssrA tag an excellent tail sequence for unfolding. In comparison, the first 11 a.a. of the other four substrates, GLGARSAGITH, should provide less grip, due to it having fewer bulky residues at 3-7th position with multiple glycine residues. Therefore, while the 11 a.a.-tail substrate does not share the same gripped sequence, it is possible to include this substrate for investigating the effect of substrate tail length, with the understanding that if the different sequences

could have an effect on the grip between ClpX and substrate, the predicted effect would be to make the grip better for ClpX. Together, when including the 11 a.a.-tail substrate, there are 5 substrates of varying lengths of tails that can be tested with the TIRF system (The tail sequences of the five substrates are shown in Table 3.5.1).

Table 3.5.1 Substrate C-terminal tail sequences.

Substrate Tail Length	Tail Sequence
11 a.a.	…**AANDENYALAA-COOH**
22 a.a.	…GLGARSAGITH**AANDENYALAA-COOH**
33 a.a.	…GLGARSAGITHLERPHRGLGD**AANDENYALAA-COOH**
48 a.a.	…GLGARSAGITHLERPHRGLGDISDQEAKPSTEDLGDK**AANDENYALAA-COOH**
85 a.a.	…GLGARSAGITHLERPHRGLGDISDQEAKPSTEDLGDK- GLGARSAGITHLERPHRGLGDISDQEAKPSTEDLGDK**AANDENYALAA-COOH**

Five DHFR substrates with unstructured C-terminal tails of different lengths. The ssrA tag is highlighted in bold. Except for the substrate that does not have a spacer between the folded DHFR domain and the ssrA tag, all, all other substrates share the same first 11 a.a. sequence

TIRF measurements were conducted on these DHFR substrates, which were MTX-stabilized, in the absence of ClpP. As described in Section 3.3, upon MTX stabilization, ClpX is no longer able to unfold the DHFR domain. The main reason to exclude ClpP in this experiment was that the protease would trim longer tails, thereby interfering with the tail-length variable. We also suspected that ClpP might affect substrate release rate by other mechanisms, and decided to first study ClpX in isolation to test properties intrinsic to ClpX. All TIRF experiments were conducted at 30 °C with 2 mM ATP. Representative TIRF images are shown in Figure 3.5.1. Time series were collected at each field of view for 10 minutes at 2 seconds intervals.

The dwell distributions of these images were then extracted using the procedure described in Chapter 2. As a reminder, puncta that lasted only a single frame were excluded in the distribution, since such a signal could be caused by non-specific binding or optical impurities. The resulting distributions are shown in Figure 3.5.2, where the total areas of the bars add up to 1. It needs to be emphasized that this normalization is slightly biased. The source of the bias is the filtered-out portion of the data, primarily the single frame traces, which differs in relative fractions for substrates with different dwell time averages. For example, for constructs with shorter average dwell time, more single-frame traces in proportion to the total traces would be excluded, making the resulting normalized

histogram skewed to the right. Also, as described in Chapter 2, puncta that were present in the first frame of the time course movie, and/or were still present at the last frame were excluded from the distribution. This might slightly bias against longer traces, but the distribution showed that the mean lifetime was mainly dictated by short interactions under 1 minute.

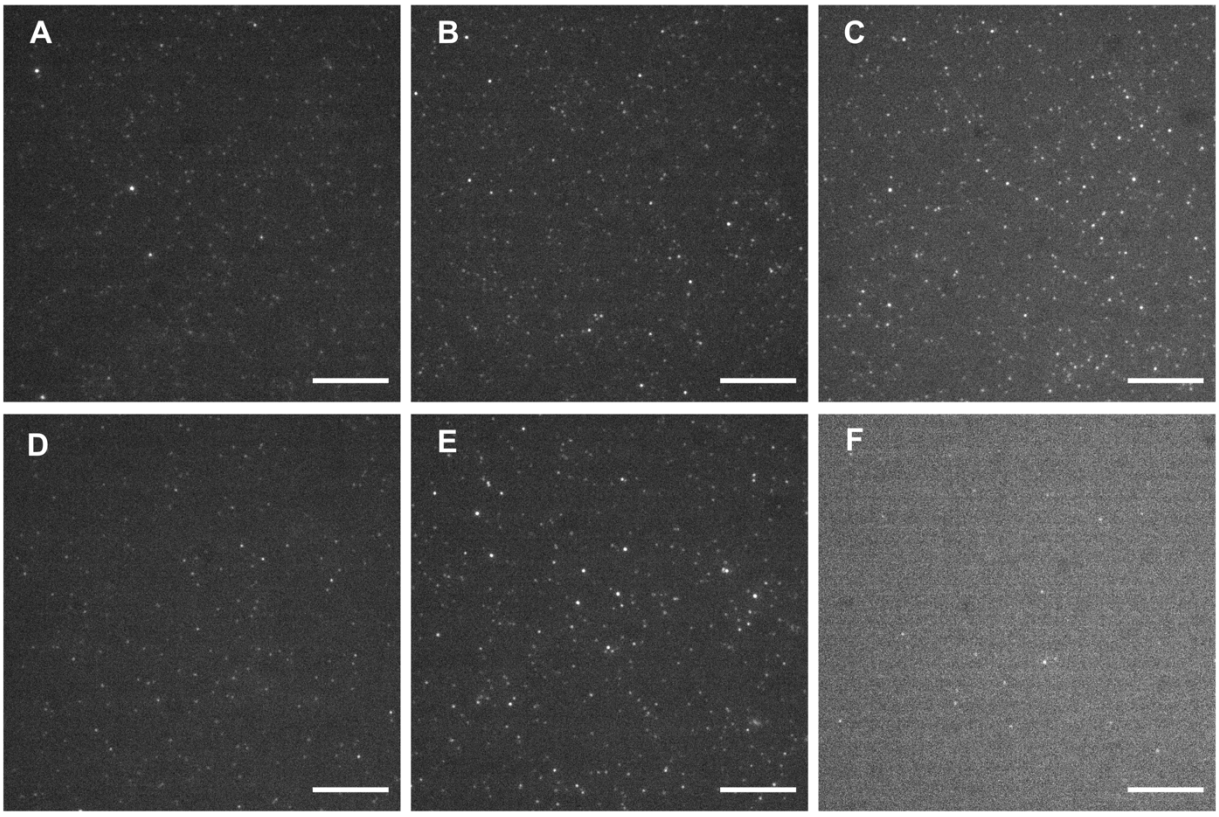


Figure 3.5.1 Representative TIRF images for measuring dwell time of substrates with different tail length.
A-E: Representative Images for tail lengths of 11, 22, 33, 48, 85 a.a. respectively. Scale bar shows 5 μm. Puncta density was kept to be around 0.04-0.1 per μm^2, or around 100 to 300 per frame at the resolution used for all imaging experiments. The puncta density of the ssrA channel was at least 10 times higher than that of measured from the ssrADD mutant channel.
F: representative image of the passivation control channel showing non-specific binding of the labeled ssrADD substrate on the coverslip, where 18 dots were counted.

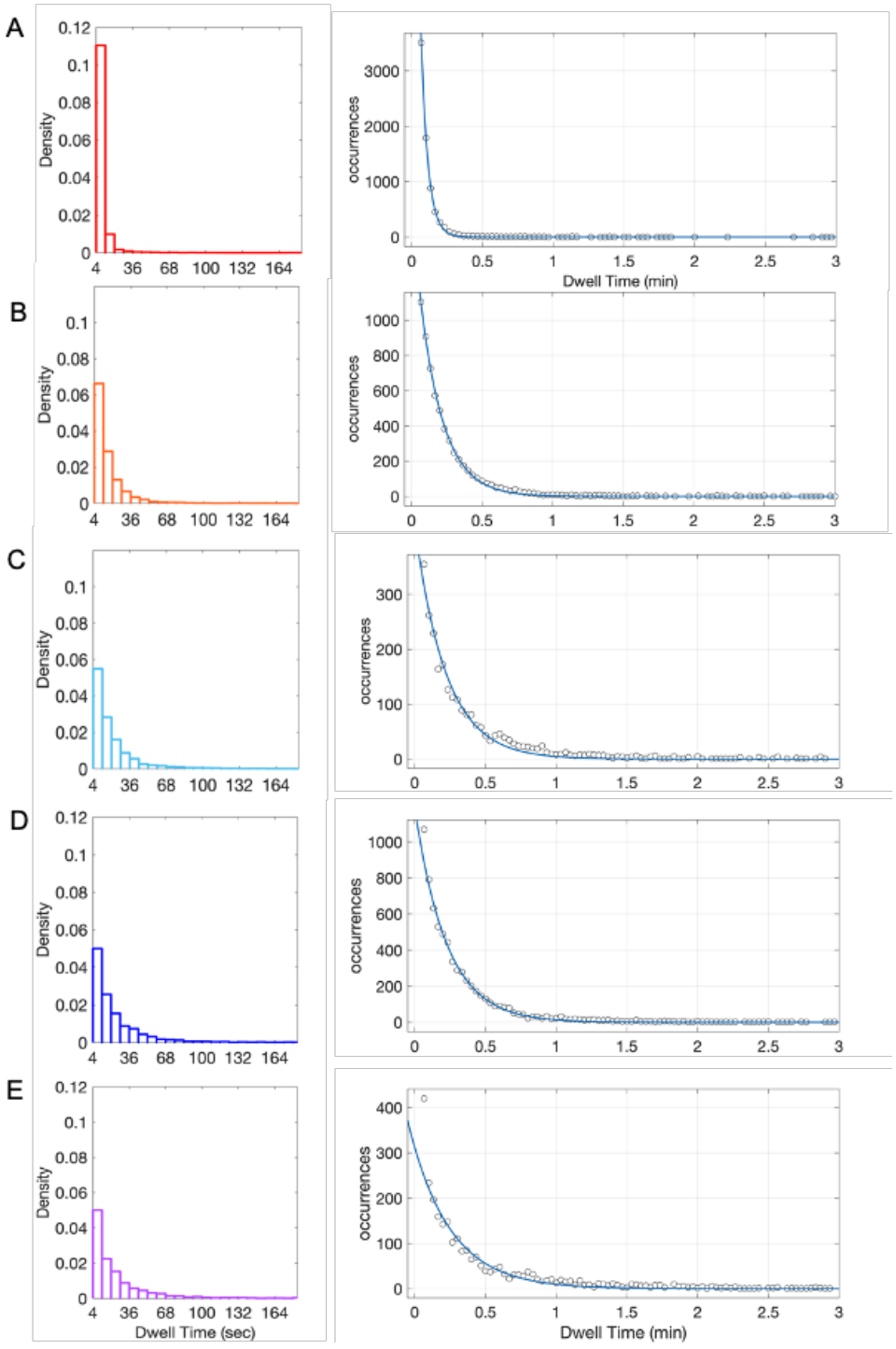


Figure 3.5.2 Dwell time distribution and curve fitting of the distributions for substrates with different tail length for their interaction with ClpX.
A to E: Normalized histograms showing dwell time distributions of substrates with 11, 22, 33, 48, 85 a.a. tails., and the corresponding curve fitting using single exponent model. The fitting was done by using model $f(t) = Ae^{-t/\tau}$, where τ is the inverse of k_{rel}. R-squared values of the curves shown were all greater than 0.99 in all five cases.

One challenge for quantitatively analyzing the data is that it is difficult to choose the appropriate underlying model for curve fitting the histogram. In this case, the exponential decay model described in **Section 3.1** is a good approximation for fitting the distributions. The mean lifetime (τ) for each substrate was obtained by curve fitting the distributions as A*e^(-t/τ).

A trend of the mean lifetime was observed, where increased tail length could also extend the mean lifetime of the ClpX-substrate complex. Without making any further assumptions for the underlying model of the dwell time distribution, the general trend was statistically compared using Kruskal-Wallis ANOVA test. The result indicated that at least one group of data was significantly different from others. A multi-way comparison among the groups was done, which showed that except for between constructs of 48 and 85 tail-lengths, all other groups were significantly different from each other (**Figure 3.5.3)**.

In **Figure 3.5.4**, the obtained mean lifetimes are plotted against the tail-lengths of the constructs. Combining the statistical comparison results in **Figure 3.5.3**, and defining $\boldsymbol{k_{rel}}$ as the inverse of mean lifetime, the data imply that tail length is a significant factor influencing $\boldsymbol{k_{rel}}$ when it is around or under 33 a.a., but its effect is no longer the dominant one after further tail extension.

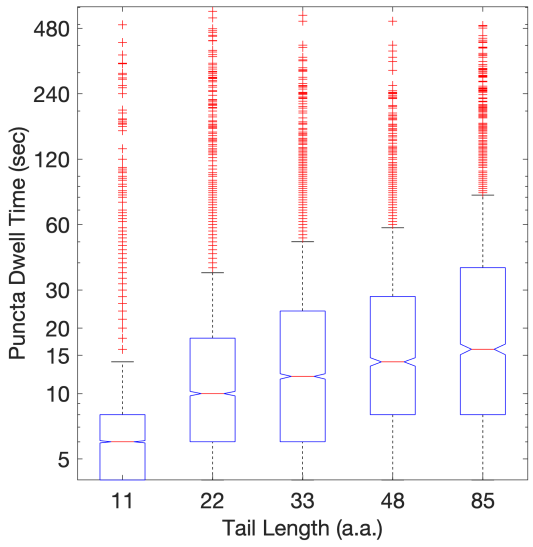


Figure 3.5.3.
Box plot showing the median and quartiles of the obtained puncta dwell time distributions for each tail length substrates. Red crosses denote statistical outliers, defined as values greater than 1.5 times away from the interquartile range.

The numbers of puncta used for each group: 7503 for 11 a.a. tail; 6244 for 22 a.a. tail; 6881 for 33 a.a. tail; 2522 for 48 a.a. tail; 2714 a.a. for 85 a.a. tail.

A multi-way Kruskal-Wallis tests showed that for the 10 pairwise comparisons, except for the one between 48 a.a. and 85 a.a. tails, all other comparisons have a $p < 0.005$ (the alpha=0.005 is set based on Bonferroni correction for 10 multi-comparisons).

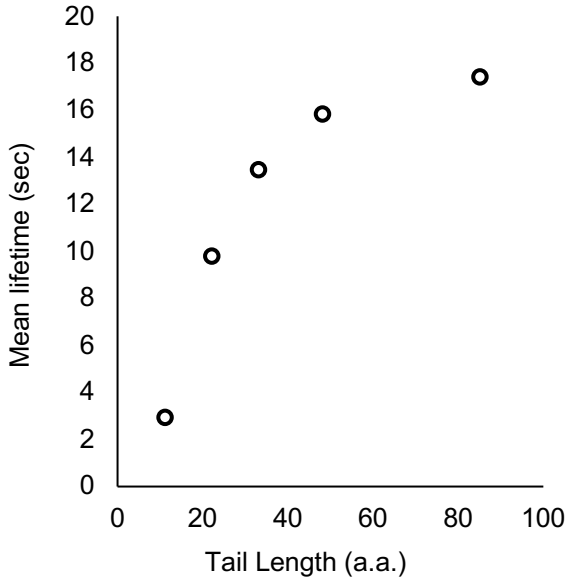


Figure 3.5.4
Mean lifetime (τ) of ClpX-substrate interactions plotted against the corresponding tail length.

Mean lifetime was obtained by fitting the dwell time distribution of each sample using with the model A*e^(-t/τ). The 95% interval of the fitted coefficient was too narrow to be properly printed.

The result supports the hypothesis that when ClpX encounters a stable domain that resists unfolding, premature substrate release is not entirely determined by events happening at the gripped region. To interpret the data, I have proposed a model where ClpX can sometimes re-establish the static interaction between tyrosine loops and the substrate after a slippage happens, as described in more detail Section 1.5**.** The proposed model predicts that for a fixed probability of losing interaction between tyrosine loops and the substrate, during the process when a substrate is backing out from the ClpX pore at certain velocity (while still being encircled by ClpX), a longer tail could afford the substrate a more prolonged temporal window to re-establish binding with the substrate.

The ClpXP cryo-EM structures indicated that the central pore of ClpX mainly interacts with the 11-12 a.a. region immediately adjacent to the folded domain. Because the inner diameter of the ClpX central pore from the top to the bottom progresses as that of an hourglass, with the narrowest point corresponding to the pore loop spiral staircase, it is unlikely that ClpX can have efficient grip on the substrate outside of the central pore. If interpreting my results in a conservative manner and thus excluding the 11 a.a.-tail substrate, all other constructs have, according to the cryo-EM structure, the same gripped polypeptide tract adjacent to DHFR. The probability of slippage over a fixed time interval at the gripped area should therefore be the same for these substrates. However, they still showed a trend of lowering release rate with increased tail length. This result therefore supports the general model that loss of static interaction between ClpX and substrate at the gripped area does not necessarily lead to the full dissociation of the substrate.

Nonetheless, the model cannot explain why dwell time cannot be further prolonged with even longer tails. The result of this limit on dwell time is that the mean lifetime of the enzyme-substrate complex is well under 1 minute, which is much shorter than the degradation rate of many stable substrates in bulk solution assays. My result argues that it could take multiple cycles of binding and release before a substrate is unfolded.

It is crucial to consider whether the surprising results of the relatively short dwell times is not due to technical artefacts. To minimize the bias against long dwell time signals, measures have been taken as described in Section 3.4 to reduce substrate background binding while improving fluorophore photo-stability. It is unlikely that photobleaching is a major concern, because the measured dwell time distributions are dominated by puncta lasting under 30 seconds, which corresponds to only 15 frames of exposure. Nonetheless, the impact of doubling the exposure was tested on dwell time distributions of a 48 a.a. tail substrate, and no statistically significant differences were observed (Figure 3.5.5). To further demonstrate the validity of the TIRF method, a SPR measurement was conducted using the 48 a.a. tail construct and monomeric ClpX, and measured the off-rate constant at 25 °C to be 0.051 s^{-1}, or a mean lifetime of 19.6 seconds (Figure 3.5.6). The result agrees with the TIRF measurements.

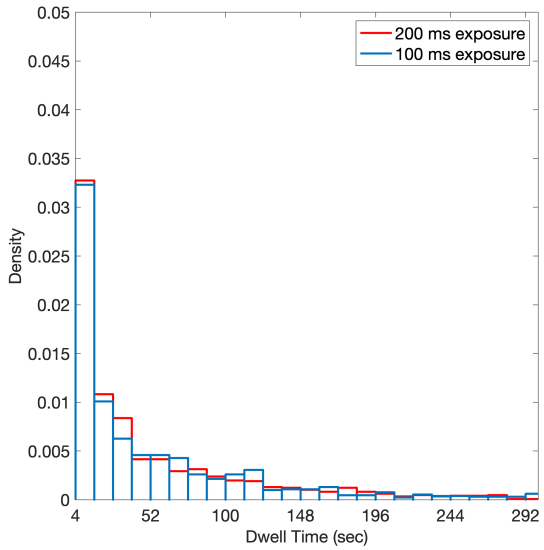


Figure 3.5.5
Normalized histogram showing the dwell time distributions of the same ClpX-substrate interaction derived from images captured at two different levels of exposures (200 ms exposure time compared to 100 ms exposure time).

Doubling the exposure time does not have a statistically significant impact on the dwell time distribution ($p = 0.21$ by Mann-Whitney U-test). In all the imaging experiments reported in this thesis, exposure was set at 100 ms.

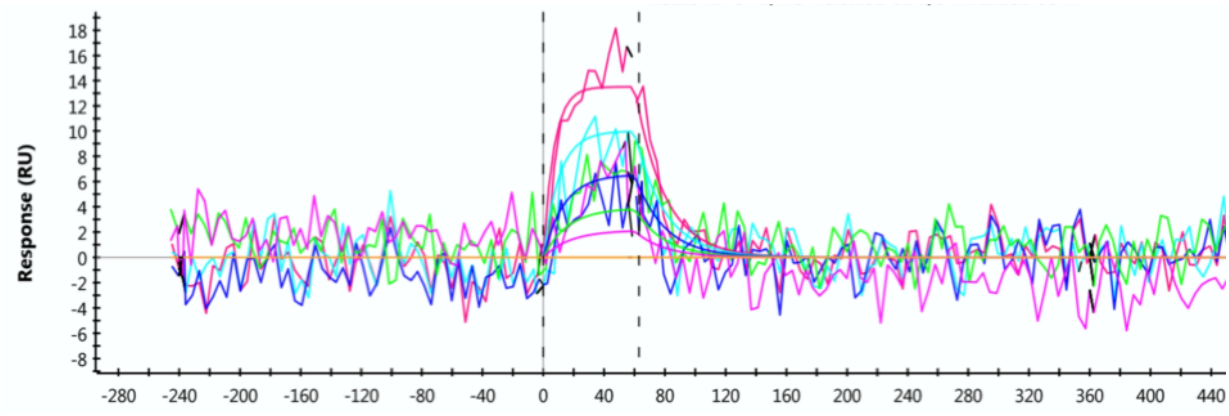


Figure 3.5.6. SPR response curves for a titration of 48 a.a. substrates. A monomeric version of the biotinylated-ClpX was immobilized on the chip; monomeric ClpX was used due to higher yield from purification. The DHFR substrate was not stabilized by MTX, due to the difficulty of having large quantity of MTX in the running buffer. Low level of degradation of the substrate was expected, but as will be shown in Figure 3.8.1, and as reported by published literature (Lee et al., 2011), degradation of DHFR by ClpXP is orders of magnitude slower than the release rate reported here, thus should not significantly affect the off rate constant. The substrates were injected at 0.25 μM, 0.5 μM, 1 μM, 2 μM, 4 μM. The off-rate constant, k_d, was fitted to 0.051 s^{-1}

In summary, the mean dwell time of substrates bound to ClpX increases with longer tails, but the effect eventually reaches a plateau where the dwell time no longer increases with increasing the length of the polypeptide. This result supports the hypothesis that the rate of release is not only determined by the gripped area. However, my model cannot explain why having longer tails above 48 a.a. does not have more effects.

3.6 Slowing down ATPase cycle prolongs ClpX-substrate lifetime

As described in Sections 1.4.3 and 1.5, the loss of static interaction between the tyrosine loops and substrate is described as the consequence of the excessive force exerted by tyrosine loops that surpasses the static friction threshold.

To briefly elaborate the model, the conclusions from structural studies introduced in Section 1.2 and the mechanochemical coupling principle introduced in Section 1.2 show that ClpX pulls the substrate by a hand-over-hand motion, and each pull is tied with the sequential hydrolysis of ATP. From these conclusions, we can assume that each pulling action by the ClpX delivers a force from the tyrosine loops to the substrate through the static interaction between them. This static interaction breaks when the gripped area of the substrate refuses to move along with the tyrosine loops. This could happen when a non-unfoldable substrate cannot be pulled further into the central pore (as the case when the substrate is already held tightly against ClpX), while, at the same time, the pulling force is higher than the threshold for maintaining the static interaction.

One prediction from this hypothesis is that when ClpX stalls at a non-unfoldable substrate, the frequency of breaking the interaction between pore loops and substrate can be reduced if the frequency of tyrosine-loop movements is also reduced.

We decided to test this hypothesis by analyzing how slowing down the ClpX ATPase cycle with ATPγS would affect the release rate of the non-unfoldable substrate, which in this case is the MTX-stabilized DHFR substrate. We predict that fewer cycles of pulling would make the stalled substrate less likely to be released.

The first to be explored was the effect of 2 mM ATPγS in the absence of ATP on substrate release. In this experiment, the lifetime of the 48 a.a.-tail substrate bound to ClpXP was measured by TIRF at 30 °C in 2 mM ATPγS. As shown in Figure 3.6.1, compared to an experiment that used the same substrate and followed the same procedure, puncta density of the same substrate (here measured as puncta per frame), averaged over 299 frames of images, was reduced 10 fold from 104 ± 12 to 9.25 ± 2 when using 2 mM ATPγS instead of 2 mM ATP. The mean dwell time from these samples was fitted to be 4.4 seconds, which is about 3-fold shorter compared to the measurement under the same condition but with 2 mM ATP (~15 seconds). The TIRF assay was not designed to quantitatively measure binding rate, because the precise amount of ClpX on the surface was not trackable, but the reduction in puncta density could not be explained by reduction in dwell time alone. This implies that substrate binding could also be inhibited by the perturbation. I did not attempt to repeat the experiment with ClpX in the absence of ClpP, due to predicted difficulty in gathering enough trace measurements and the uncertainty about the binding state of the substrate.

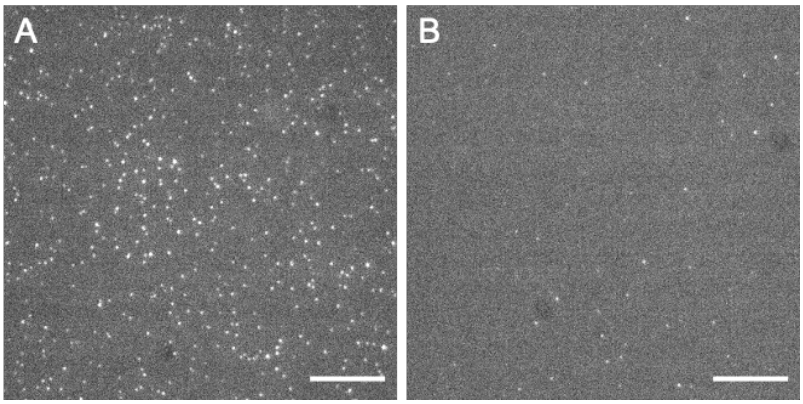


Figure 3.6.1. Comparing to the same substrate at 2 mM ATP in A, puncta density is apparently reduced when 2 mM ATPγS is the only source of energy for ClpXP as is shown in B. Puncta density, averaged over the 299 frames of images was reduced from 104 ± 12 to 9.25 ± 2 in the presence of ATPγS.

The apparent hypothesis to explain the reduced dwell time under 2 mM ATPγS without ATP is that the substrate tail was not fully inserted into the central pore due to reduced translocation speed. It has been proposed that substrate binding to ClpX is an ATP dependent process as a result of a two-step binding mechanism (Martin et al., 2008b). Because the initial substrate binding to the central pore is equivalent to a brief translocation stage, significant reduction in translocation speed would have an impact on the efficiency of this process. Translocation speed has been measured before using optical tweezers, and it was shown that it becomes slower when increasing concentrations of ATPγS are added to a constant background of ATP; in pure ATPγS, translocation is almost halted (Sen et al., 2013). Because ClpX hydrolyzes ATPγS at a rate 20 to 40-fold slower compared to that of ATP hydrolysis (Sen et al., 2013), it is much less likely for substrate tails to be sufficiently seated when ATPγS is the only energy source, as in my experiment. Based on the result in Section 3.5, such condition would be similar to having a short tail for the substrate, making it more likely to be released. Apparently, the blunt method of perturbation by 2mM ATPγS was not an ideal strategy to test the stated hypothesis, although the result does provide certain hypotheses on ClpX functions.

To circumvent the substrate binding problem, a mixture of ATPγS and ATP was used, which would slow down ATPase activity to impair unfolding, but still fast enough to allow ClpX to reach and stall at the folded domain. As introduced in **Section 1.2.2.,** the effect of ATPγS, when used as a competitive inhibitor for ATP, is that the binding of ATPγS pauses the ATPase cycle. Sen et al. (2013) showed in their optical tweezers studies that ATPγS inhibition is limited by its off-rate instead of the rate of its hydrolysis under this

condition; once ATPγS is released, ClpX should resume its normal rate of ATPase activity. Therefore, the more ATPγS is bound per ClpX, the less likely it is for ClpX to resume normal ATPase activity.

It needs to be emphasized that inhibition of ClpX ATPase activity by ATPγS is not equivalent to using lower concentrations of ATP. This is best demonstrated by the results from Sen et al. (2013). When measuring ClpX step sizes with optical tweezers, lowering ATP concentration does not change the frequency of ClpX mechanical cycle, only reducing the maximum number of steps ClpX can take in a burst. In comparison, ATPγS binding induces long pauses in ClpX mechanical cycle. One consequence is that ClpX translocation velocity is not sensitive to changes in ATP concentration, but it is very sensitive to the presence of ATPγS.

To find this optimal ATPγS/ATP ratio for TIRF imaging, we did two bulk-solution assays to test the effect of different ATPγS-to-ATP ratios on ClpX activity. We first assayed the effect of ATPγS on the degradation activity of ClpXP (which, being a degradation assay, naturally requires the presence of ClpP). Here, cpGFP, an easy to unfold substrate compared to DHFR, was used as the benchmark substrate for ClpXP degradation activity, and assessed the inhibition of ClpXP by various concentrations of ATPγS in a constant background of 2 mM ATP. As reported by Nager et al. (2011), cpGFP can be degraded by ClpXP by as low as 50 μM ATP (Km for ATP has been reported to be 59 μM by Sen et al.), and its fractional degradation rate is roughly linearly increased with increasing ATPase rate. Therefore, this substrate is suitable to function as the benchmark for finding the lowest ATPγS level that still supports degradation of native substrates. In this assay, ClpXP retained about 10% degradation activity of cpGFP when ATPγS was raised to 250 μM (Figure 3.6.2).

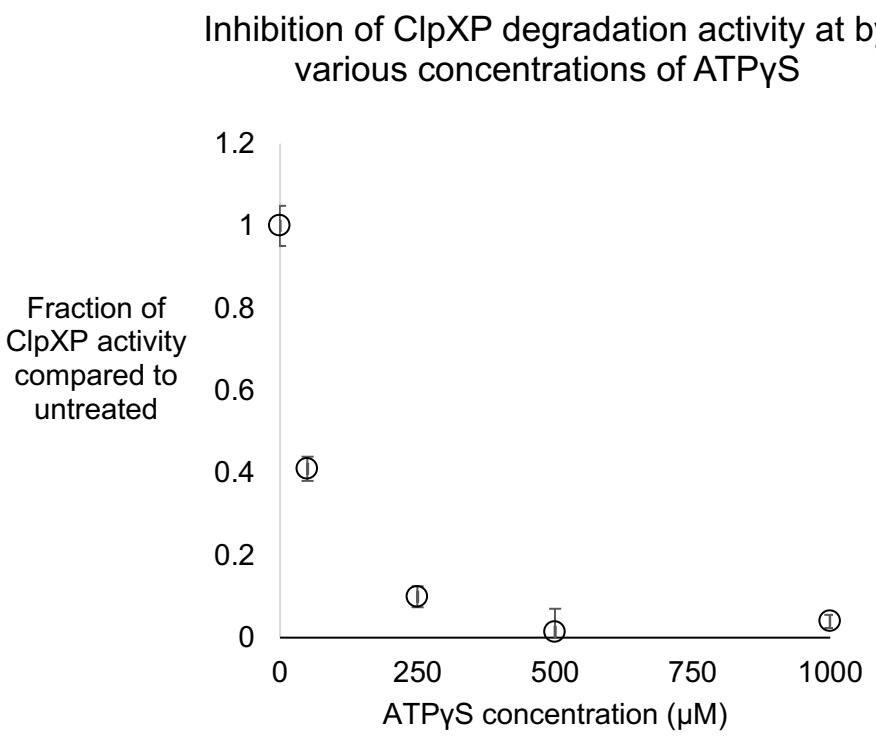


Figure 3.6.2. Degradation activity of ClpXP at 2 mM ATP, inhibited by increasing level of ATPγS. At 250 μM ATPγS, ClpX still retains about 10% degradation activity, suggesting that substrate binding is not abolished.

We also wanted to know the ATPase rate of ClpX (and ClpXP) at this condition. In this NADH-coupled ATPase assay, to imitate the TIRF experiment setting, I fed the ClpX and ClpXP with 10 μM of the 48 a.a. Tail DHFR substrate without MTX stabilization. The reason to leave out MTX was that its spectrum has some overlap with that of NADH, which would reduce the sensitivity of the measurement.

A side note for this NADH-coupled ATPase assay is that it measures the sum of hydrolysis events for both ATP and ATPγS, but the latter is not likely to be hydrolyzed during its binding with ClpX. Here, the rates were reported simply as “hydrolysis per minute”.

As shown in Figure 3.6.3, for ClpX, the ATPase rate is reduced from 117 ± 7.2 hydrolysis per minute to 84 ± 7.6 hydrolysis per minute upon ATPγS inhibition at 250 μM; in the presence of ClpP, the ClpXP ATPase rate was reduced from 79.9 ± 5.1 to 62.9 ± 3.22 per minute upon ATPγS inhibition. The result indicates that the extent of ClpX activity reduction by 250 μM ATPγS should be minor enough to not dramatically impair substrate binding at a populational level.

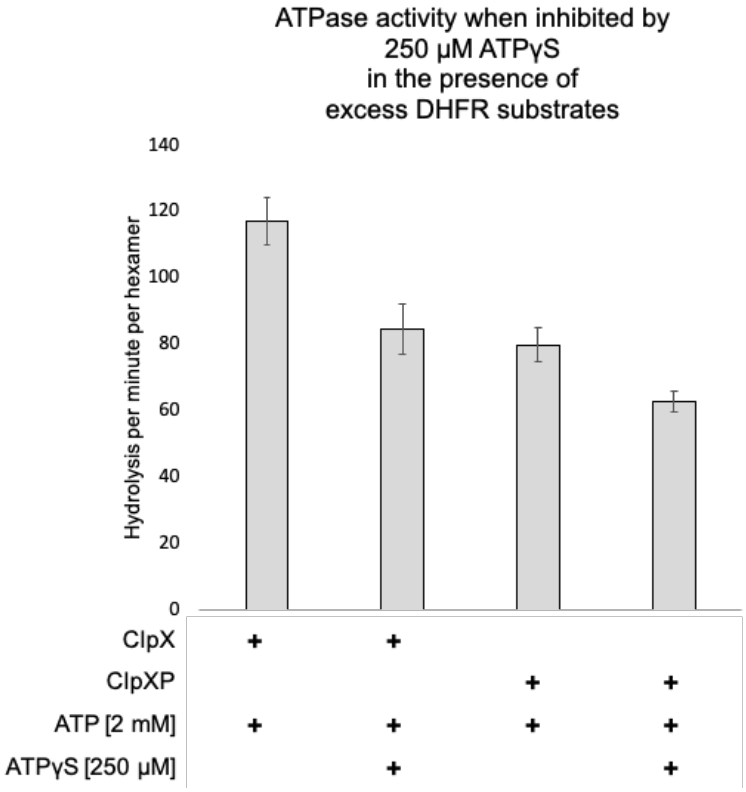


Figure 3.6.3 ATPase activity of ClpX in the presence of the DHFR substrate. ATPγS at 250 μM in combination with 2 mM ATP reduces ClpX ATPase activity from 117 ± 7.2 hydrolysis per minute to 84 ± 7.6 per minute. For ClpXP, rate was reduced from 79.9 ± 5.1 to 62.9 ± 3.22

As stated at the beginning of this section, my hypothesis requires the substrate to be held tightly against ClpX; for a substrate with a long tail, pausing ClpX ATPase activity by ATPγS has a higher probability of trapping ClpX at the tail region before it reaches the folded domain. To avoid this potential problem, the DHFR substrate with 11 a.a.-tail was used for TIRF measurements, which should only take fewer than 2 steps for ClpX to translocate the entire tail region, based on published optical tweezers studies (Abuin-Tam et al., 2011; Maillard et al., 2011).

When examined with TIRF microscopy, as shown in Figure 3.6.4, addition of 250 μM ATPγS did not qualitatively impair substrate binding, compared to 2mM ATP alone, as reflected in the similar puncta density (87.2 ± 11 per frame) compared to the control without 250 μM ATPγS (95.2 ± 12 per frame) . When compared to its mean lifetime under the control condition (3.1 ± 0.01 sec), the substrate had a longer mean lifetime (4.9 ± 0.05 sec) in the presence of 250 μM ATPγS (Figure 3.6.5).

Because ClpX ATPase activity can be reduced upon binding to ClpP (Joshi et al., 2004), substrate dwell time was also measured using ClpXP. ClpP binding increased substrate dwell time to 6.0 ± 0.03 secs, while ATPγS further prolonged the dwell time to 7.4 ± 0.06 secs (Figure 3.6.6). It should be emphasized that although the mechanism by which ClpP reduces ClpX ATPase activity is unknown, it is established that unlike ATPγS, binding to ClpP does not induce pauses in ClpX ATPase cycle, as shown in optical tweezers studies (Abuin-Tam et al., 2011; Maillard et al., 2011).

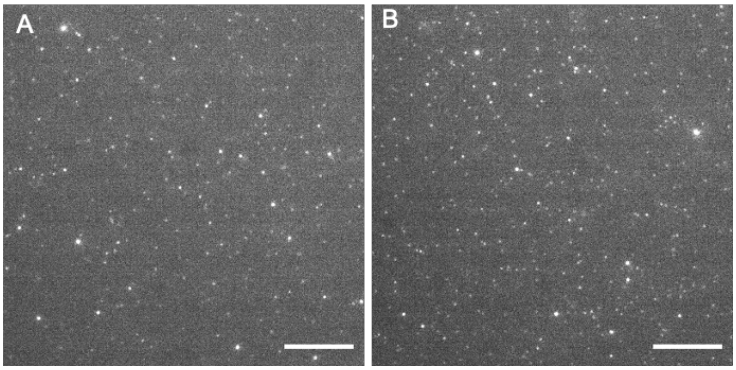


Figure 3.6.4. Comparing to the same substrate at 2 mM ATP in (A), puncta density (95.2 ± 12 per frame) is not reduced when in B), 250 μM ATPγS is present (87.2 ± 11 per frame)

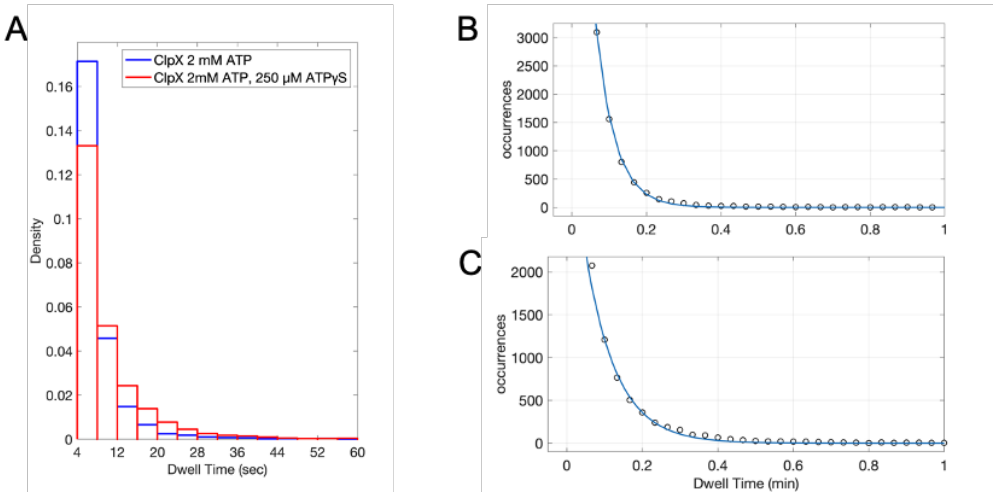


Figure 3.6.5. Dwell time distributions and curve fittings of the distributions for the 11 a.a.-tail substrate binding with ClpX (without ClpP**)** in the absence or in the presence of 250 μM ATPγS.
A). Normalized histogram overlaying dwell times in the absence of ATPγS (blue) or in the presence of 250 μM ATPγS (red).
B). Curve fitting for the lifetime distribution of the control group. Mean lifetime = 3.1 ± 0.01 sec; R-squared > 0.99.
C). Curve fitting for the lifetime distribution of the group treated with 250 μM ATPγS. Mean lifetime = 4.9 ± 0.05 sec; R-squared > 0.99

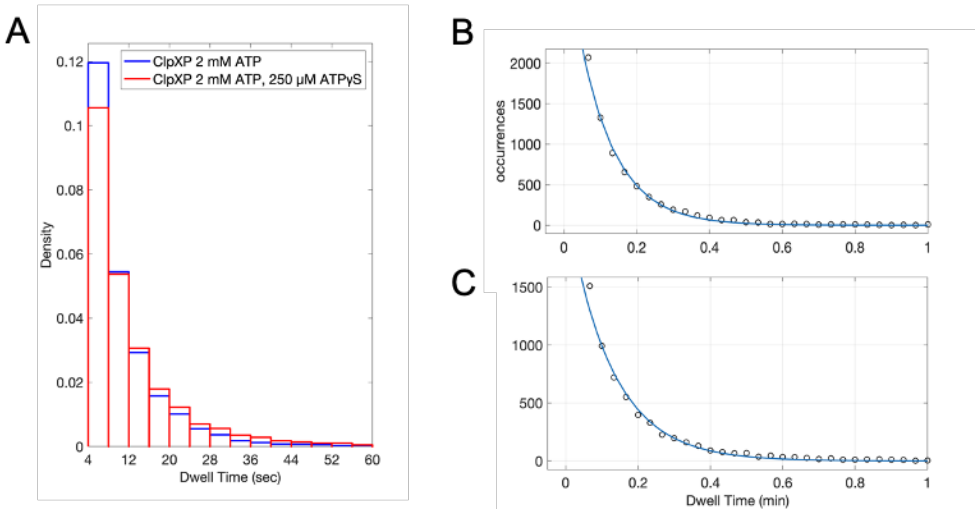


Figure 3.6.6 Dwell time distributions and curve fittings of the distributions for the 11 a.a.-tail substrate binding with **ClpXP** in the absence or in the presence of 250 μM ATPγS.
A). Normalized histogram overlaying dwell times for ClpXP in the absence of ATPγS (blue) or in the presence of 250 μM ATPγS (red) for.
B). Curve fitting for untreated population. Mean lifetime = 6.00 ± 0.03 seconds; R-squared > 0.99
C). Curve fitting for 250 μM treated population. Mean lifetime = 7.41 ± 0.07 seconds ; R-squared > 0.99

This result showed that when substrate is stalled at ClpX, there exists a low affinity step within the ClpX mechanical cycle, such that slowing down its conformational change reduces the frequency of the low affinity step, thus reducing the overall k_{rel}.

The effect of ClpP is more difficult to interpret, because the full effects of ClpX-ClpP interaction on the dynamics of both particles are not thoroughly understood. However, the reduced ClpX ATPase activity upon ClpP binding means that the kinetic effects of ClpP association could be one of the many mechanisms that ClpP use to improve substrate processivity.

In summary, the results showed that reducing ATP activity may increase substrate dwell time when the substrate is held against ClpX. While this result supports the prediction of my original hypothesis that reducing conformational change frequency would in turn reduce the chance to break the static force interaction, proving this hypothesis still requires force measurements, which cannot be achieved with my current TIRF system.

3.7 ClpP binding to ClpX improves ClpX-substrate lifetime

In the previous section, it was briefly touched upon that ClpP binding to ClpX could potentially reduce k_{rel} by slowing down ClpX ATP hydrolysis. However, this hypothesis does not exclude other mechanisms by which ClpP could affect k_{rel}. The experiment mentioned in Section 3.6 used a short tail substrate, which was not likely to be translocated deeply enough to have any interaction with ClpP as demonstrated by structures from Ripstein et al. (2020). It is possible that having a longer tail that can reach into ClpP could further reduce k_{rel}. A prediction from this hypothesis is that gradually increasing tail length could dramatically change substrate dwell time, in a manner dependent on tail access to ClpP.

In this experiment, we measured the dwell time distribution of MTX-stabilized DHFR substrates of different tail lengths bound to ClpXP at 2 mM ATP. The substrates were the same ones used in Section 3.5. In the presence of ClpP, all substrates with tails longer than 11 a.a. fitted relatively poorly with a single exponential decay model (Figure 3.7.1). By examining the curve fitting result, it was apparent that a slower decay process was unaccounted for by the single exponent model, potentially resulting in the fitted parameter underestimating the true mean lifetime. This is in sharp contrast with the experiments in the absence of ClpP. This indicates that there potentially exists a second population of ClpXP-substrate complexes with a much longer mean lifetime on the order of minutes.

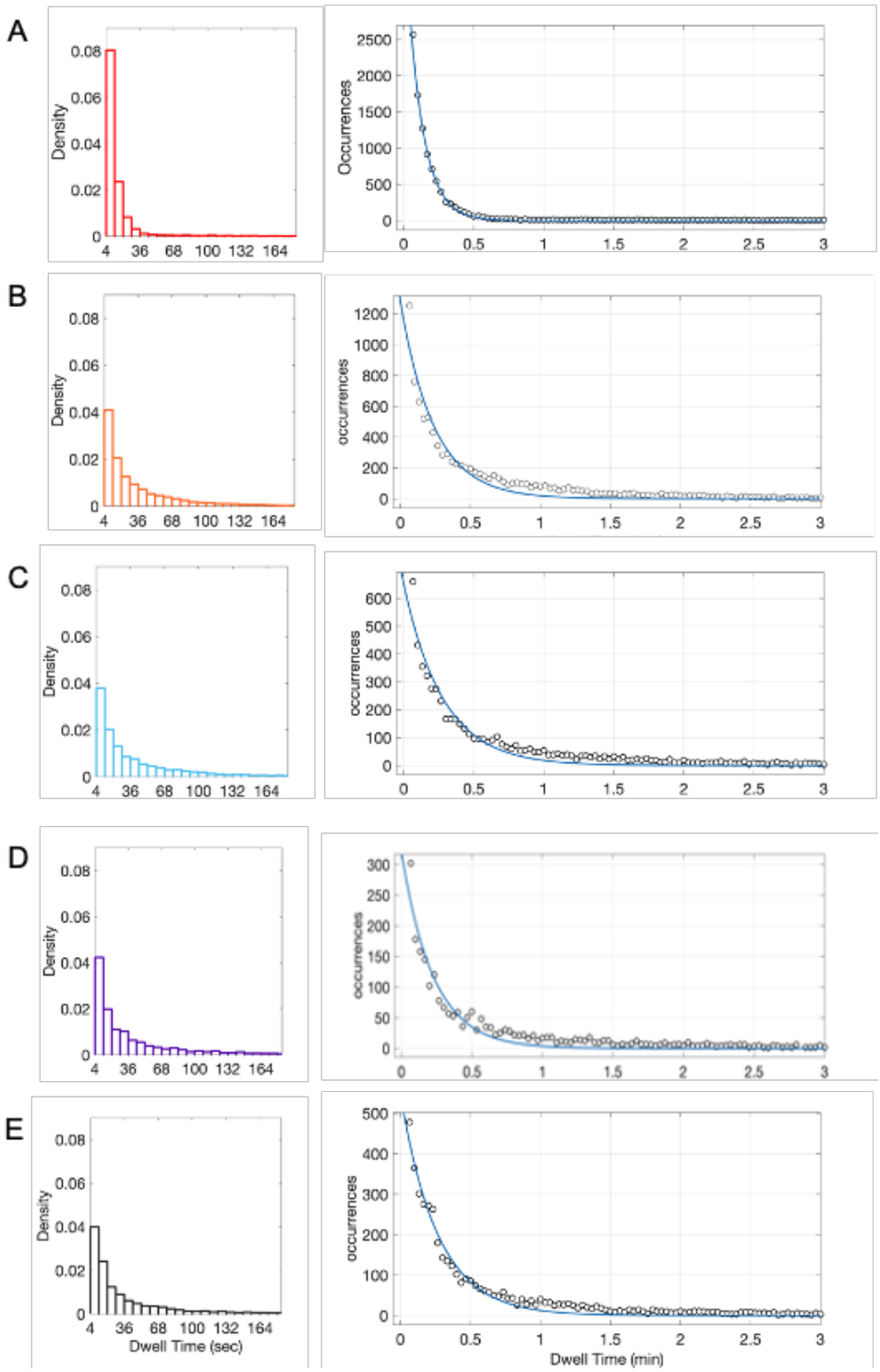
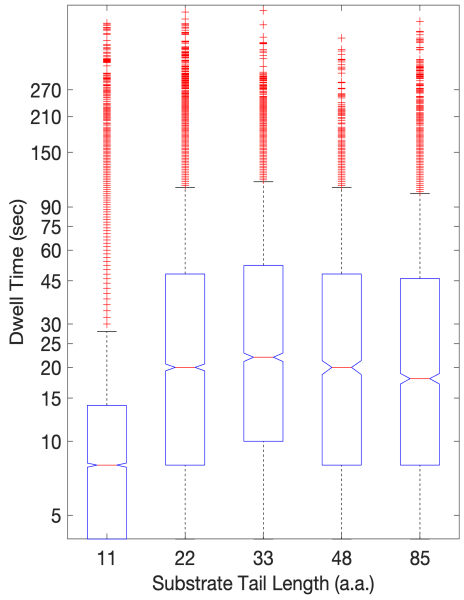


Figure 3.7.1 Dwell time distribution and curve fitting of the distributions for substrates with different tail length for their interaction with ClpX.
A to E: Normalized histograms showing dwell time distributions of substrates with 11, 22, 33, 48 and 85 a.a.-tails interacting with ClpXP. Note that the single exponential decay model does not satisfactorily describe the distributions for substrates in B to E. The fitting was still done by using model $f(t) = Ae^{-t/\tau}$. R-squared values for distributions shown from A to E are 0.997, 0.937, 0.945, 0.927, 0.972 respectively.

A



B

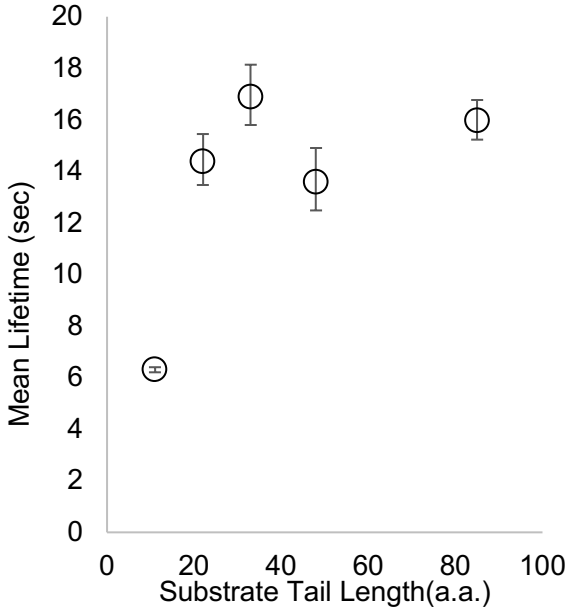


Figure 3.7.2

A) Box plot showing the median and quartiles of the measured puncta dwell times for each tail length. Red crosses denote statistical outliers, defined as values greater than 1.5 times away from the interquartile range.

Counts for each group: 11 a.a.-tail, n=10085; 22 a.a. tail, n=9644; 33 a.a. tail, n= 5831; 48 a.a. tail, n = 2313; 85 a.a.-tail, n=4447.

A multi-way Kruskal-Wallis tests showed that for the 10 pairwise comparisons, only the 11 a.a.-tail substrate has a population that is statistically different from other populations (P<0.005; alpha is set to 0.005 for multiple comparisons)

B) Fitted mean lifetime (τ)of ClpXP-substrate interactions plotted against the corresponding tail length. Mean lifetime was obtained by fitting the distribution with $\mathrm{f(t)} = Ae^{-t/\tau}$. Error bar represents the 95% confidence interval of the fitted coefficient. When alpha is set to 0.005, only the 11 a.a.-tail substrate has a fitted mean lifetime significantly different from the other four substrates.

However, a couple of key parameters for setting the boundaries of the two exponent models are unknown. The only values that could be used to restrict the model were the decay rates obtained from measurements of the same substrates in the absence of ClpP, as described in Section 3.5, but it would be problematic to do so. If it were assumed that the fast component could be fitted using this decay rate, it would indicate that this fast component is ClpP independent. We do not have evidence to support this assumption. Therefore, for this set of data, the focus is on establishing a trend, while avoiding over-interpreting the absolute value of the fitted decay rate from the single exponent model.

When looking at the nature of the relationship between tail length and duration of dwell time in the presence of ClpP, it appears that the curve plateaus faster compared to that obtained in the absence of ClpP, as shown in Section 3.5 (Figure 3.7.2). However, it needs to be stated once again that the 11 a.a.-tail substrate has a different gripped sequence by ClpX compared to other substrates. As discussed in Section 3.5, the grip is not likely significantly changed, but nonetheless the trend shown in Figure 3.7.2 should be concluded with reservations.

Judging from the published structures, it is likely that 22 a.a.-tail could reach ClpP, but its contact need be confirmed, either with structural methods or with crosslinking methods. If this assumption can be supported by future experiments, it can be concluded the off-rate is no longer dominated by tail length if the tail could reach into ClpP. This observation will then support the conclusion that possible interactions between substrate and ClpP could help reduce k_{rel}.

However, it is not fully understood how ClpP affects ClpX dynamics. Because of the ability of ClpP to modulate ClpX ATPase cycle, it is reasonable to expect that ClpP could change ClpX dynamics. Due to this possibility, the above TIRF result cannot refute the alternative hypothesis where due to changed dynamics, ClpX is much more efficient at re-establishing static interaction between pore loops and substrates, thus effectively reducing the need to have a longer tail. Again, further studies are needed to distinguish these two alternative hypotheses.

3.8 A single-molecule level confirmation that glycine-alanine repeats do not cause substrates to have a higher release rate

On the topic of the effect of glycine-alanine rich sequence (which was discussed in more detail in Section 1.4.4), there was a debate between Priscilla Too et al. (2013) and Daniel Kraut on whether a substrate with glycine alanine repeats (GAr) had higher k_{rel} from ClpXP. The resolution of the debate was that this specific GAr substrate has largely the same k_{rel} as a control substrate, and that the accumulation of the unprocessed fragment was caused by a significant reduction in the rate of the forward reaction. Because the measurements by Too et al. were done in a bulk solution experiment, the TIRF assay described in this thesis could contribute to the conclusion by making the comparison at the single molecule level.

Too et al. reported that ClpXP releases a higher amount of unprocessed titin-I27 domain when a 10 a.a. GAr is placed 22 a.a. away from the C-terminal end of this domain; shortening the spacer between the GAr and titin-I27 domain decreases the amount of release products. In the TIRF experiment described in the following, a GAr sequence was similarly placed at 22 a.a. from the C-terminal end of the cpGFP-DHFR substrate. The tail of this DHFR-GAr substrate was constituted of a 22 amino acid spacer, a 15 amino acid GAr sequence, and the ssrA tag. To simplify nomenclature, this substrate is referred to as the GAr15 substrate. The only difference between GAr15 and the 48 a.a.-tail substrates mentioned in Sections 3.5-3.7 was the 15 a.a. glycine alanine repeat (test sequences used in this section is listed in Table 3.8.1). Thus, the original 48 a.a-tail substrate without the GA repeats is referred to as the control substrate in this section.

Table 3.8.1 Tail sequences for 11 a.a. tail substrate, the 48 a.a. tail control, GAr15, and GA37.

Substrate	Tail Sequence
11 a.a.	…**AANDENYALAA-COOH**
48 a.a. Control	…GLGARSAGITHLERPHRGLGDISDQEAKPSTEDLGDK**AANDENYALAA-COOH**
GAr15	…GLGARSAGITHLERPHRGLGDIAGAGGGAGAGGAGGA**AANDENYALAA-COOH**
GAr37	…AGAGGAAGAGGGGAAGAAAAAGAGAGAGGAGAAAGGG**AANDENYALAA-COOH**

The ssrA tag is bolded.
The glycine-alanine repeats are labeled in red

It needs to be noted that the kinetic model for the degradation of the TIRF substrate by ClpXP differs from the ones used in Too et al. (2013). Specifically, when not stabilized by MTX, in the TIRF substrate the DHFR domain is degraded before its cpGFP domain. The kinetic model of this reaction at steady state is:

$$E + S \underset{k_{rel}}{\overset{k_1}{\rightleftharpoons}} ES_1 \xrightarrow{k_2} ES_2 \xrightarrow{k_3} E + P$$

The following assumptions were made for the kinetic model. First, the premature release, k_{rel} is simplified to be the reverse rate constant of the first step reaction, k_{-1}. Second, the premature release of the cpGFP domain is not considered, because the cpGFP has been shown to be degraded by ClpXP at low ATP concentration, suggesting a much lower energy threshold for unfolding, which typically results in increased processivity compared to the DHFR domain (Nager et al., 2011; Bell et al., 2019). When using this model, the degradation rate measured by tracking the loss fluorescence of the GFP domain becomes

$$V = k_3[ES_2] = e_0 \frac{a_1 k_1 k_2 k_3}{k_{-1}k_3 + k_2 k_3 + a_1 k_1 k_2 + a_1 k_1 k_3} = e_0 \frac{a_1 k_1}{\frac{k_{-1} + a_1 k_1}{k_2} + \frac{k_3 + a_1 k_1}{k_3}}$$

At fixed concentrations of substrate and ClpXP, when assuming that $\boldsymbol{e_0}$ (protein concentration)**,** $\boldsymbol{a_1}$ (substrate initial concentration)**,** $\boldsymbol{k_1}$**,** $\boldsymbol{k_2}$**,** $\boldsymbol{k_3}$ are the not changed for substrates of different tail sequence, the changes in reaction rate is decided only by $\boldsymbol{k_2}$ (degradation rate of the DHFR domain) and $\boldsymbol{k_{-1}}$ (the release rate of the unprocessed substrate). However, hypothetically speaking, $\boldsymbol{k_2}$ is very small for the DHFR substrate, which would make the rate *V* not sensitive to further reduction of $\boldsymbol{k_2}$ in the model above. As mentioned in **Section 3.3**, this was a compromise of using the substrate designed TIRF assay to be used directly for bulk solution measurements degradation. While substrates for this work served the purpose of simplifying the kinetic model for the TIRF experiment, it should be acknowledged that alternative substrates are preferred for the purpose of bulk solution measurements, which, unfortunately, were not used in this thesis.

In a fluorescence based bulk solution degradation assay, under the above kinetic scheme, measurements were done for degradation of 2.5 μM of control, GAr15, the 11 a.a.-tail substrates, as well as another substrate with a more extensive GAr tract (GAr37) which is to be described shortly. As shown in Figure 3.8.1, all DHFR substrates have low degradation rates even without MTX stabilization. This low degradation rate had made it difficult to observe the differences between substrates. With these limitations, the

degradation rate was measured to be 12% slower for the GAr15 when compared to the control; to put this into perspective, the 11 a.a.-tail substrate was degraded 45% slower than the control; in the presence of MTX, no significant degradation was detected for the control.

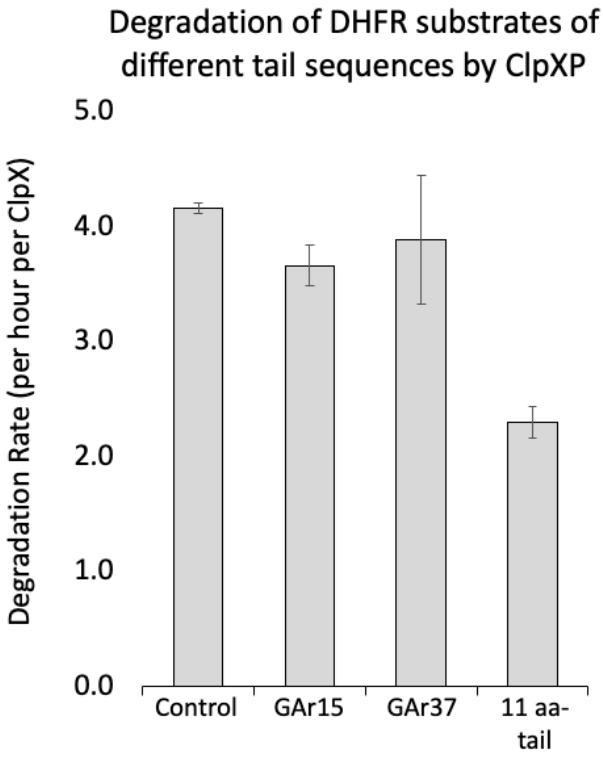


Figure 3.8.1 Comparison of degradation rates of DHFR substrates (2.5 μM) with different tail features by ClpXP at 2 mM ATP. Degradation assay was conducted using the loss of fluorescence of the cpGFP domain. Rate was calculated by the fraction of fluorescence lost over duration t. Degradation rate of GA37 was not statistically different from GAr15 and the control due to large variance.

When measuring the dwell time distribution of under TIRF, in the absence of ClpP, the difference between the two populations were very small; mean ClpX-substrate lifetime was slightly longer at 14.4 ± 0.1 sec for GAr15, compared to 13.3 ± 0.3 for the control (Figure 3.8.2). In the presence of ClpP, both samples suffered the same curve fitting problem as described in Section 3.7 where the fitted model would potentially underestimate the mean lifetime. The approximated bounds of the mean lifetimes for GAr15 and control overlapped (14.7 ± 1.6 seconds for GAr15; 13.6 ± 1.3 seconds for the control). The result indicates that GAr15 indeed does not increase substrate release rate, as originally reported by Kraut et al. (2012 and 2013)

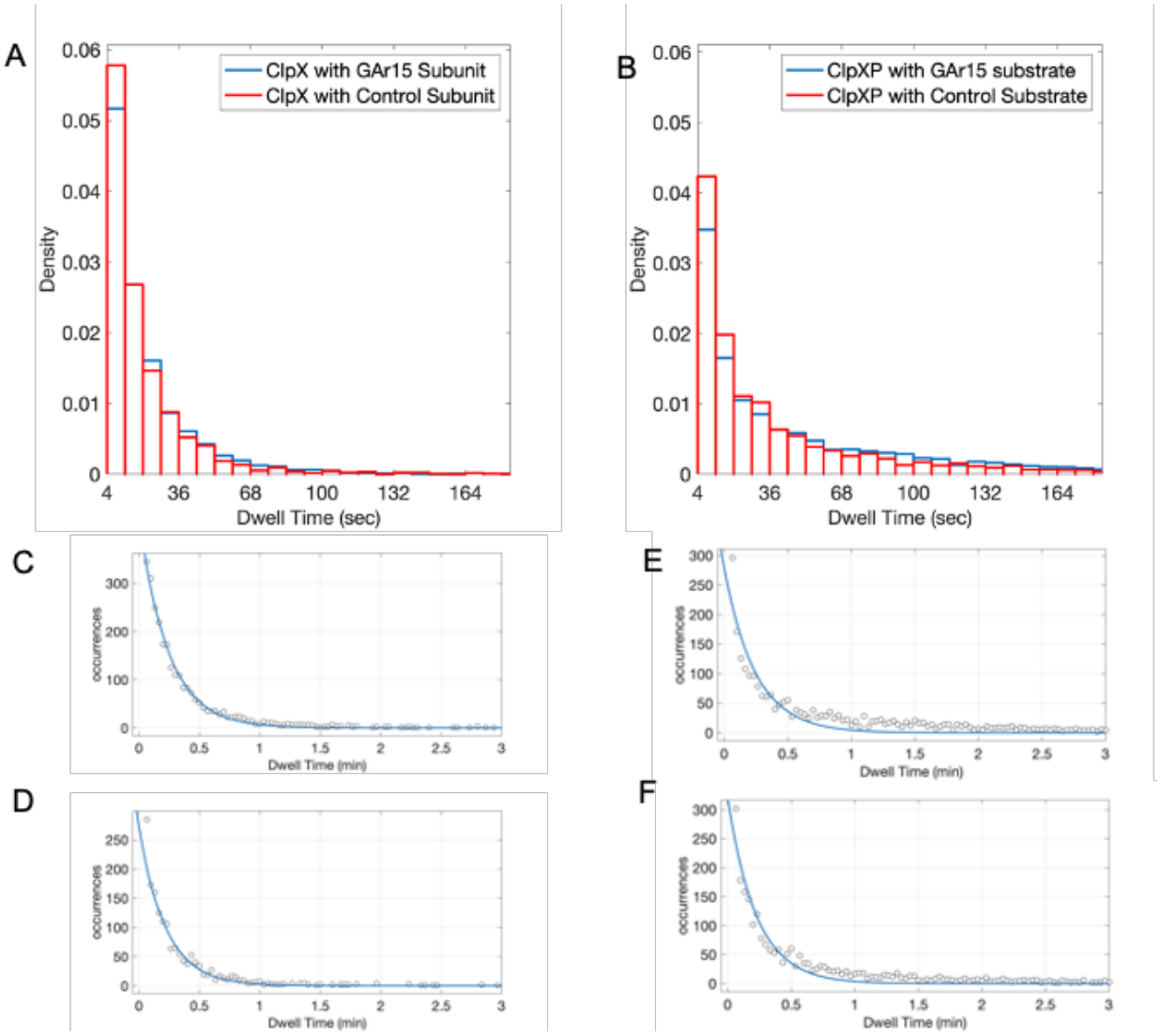


Figure 3.8.2 Dwell time distributions, and curve fittings of the distributions for the GAr15 substrate and the control substrate interacting with ClpX or ClpXP.

A) Histograms showing the overlay of normalized dwell time distributions for the GAr15 substrate (blue bars) and the control substrate (red bars) interaction with ClpX (in the absence of ClpP)

B) Histogram showing the overlay of normalized dwell time distributions for GAr15 and the control substrate interacting with ClpXP.

C – F): Curve fitting over the dwell time distributions using the single exponential decay model to extract τ.

C) Curve fitting for ClpX-GAr15 interaction; $\tau = 14.4 \pm 0.1$; R-squared > 0.99

D) Curve fitting for ClpX-Control interaction; $\tau = 13.3 \pm 0.3$; R-squared > 0.99

E) Curve fitting for ClpXP-GAr15 interaction; $\tau = 14.7 \pm 1.6$; R-squared > 0.99

F) Curve fitting for ClpXP-Control interaction; $\tau = 13.6 \pm 1.3$; R-squared =0.926

The above experiment was designed before the ClpXP cryo-EM structure became available. After studying the ClpXP structures by Fei et al. published in 2020, it was realized that the pore loops of ClpXP should grip the 22 a.a. spacer instead of the GAr15 sequence when trying to unfold the DHFR domain. This warranted another test, because the GAr region in GAr15 was not properly seated on the pore loops of ClpX. In Chapter 4, Section 4.5, I will discuss how the GAr in the GAr15 substrate could hypothetically affect substrate degradation without directly interacting with the ClpX tyrosine loops. Here, to impose the GAr sequence on ClpX pore loops, another 48 a.a.-tail substrate was produced, of which the tail was made of a 37 a.a. GAr sequence and 11 a.a. ssrA tag. The large stretch of GAr sequence in the tail region should ensure ClpX grips the glycine-alanine repeats during substrate unfolding. This substrate is referred to as GAr37.

The fluorescence-based bulk solution degradation assay was used to test the degradation rate GAr37 by ClpXP, which was measured to be not statistically different from GAr15 or the control, as already shown in Figure 3.8.1. Because the decision to measure GAr37 was close to the time of writing this thesis, I only managed to measure its dwell time distribution in the absence of ClpP. The mean lifetime of GAr37 was similar to GAr15, measured at 14.29 ± 1.1 seconds (Figure 3.8.3). However, the single exponential decay model used in Section 3.5 does not fit well with the GAr37 dwell time population. Therefore, the fitted parameter likely underestimated the true mean lifetime, as seen from the histogram overlay. Nonetheless, this result unquestionably confirms that glycine-alanine repeats do not cause substrates to be more readily released by ClpX.

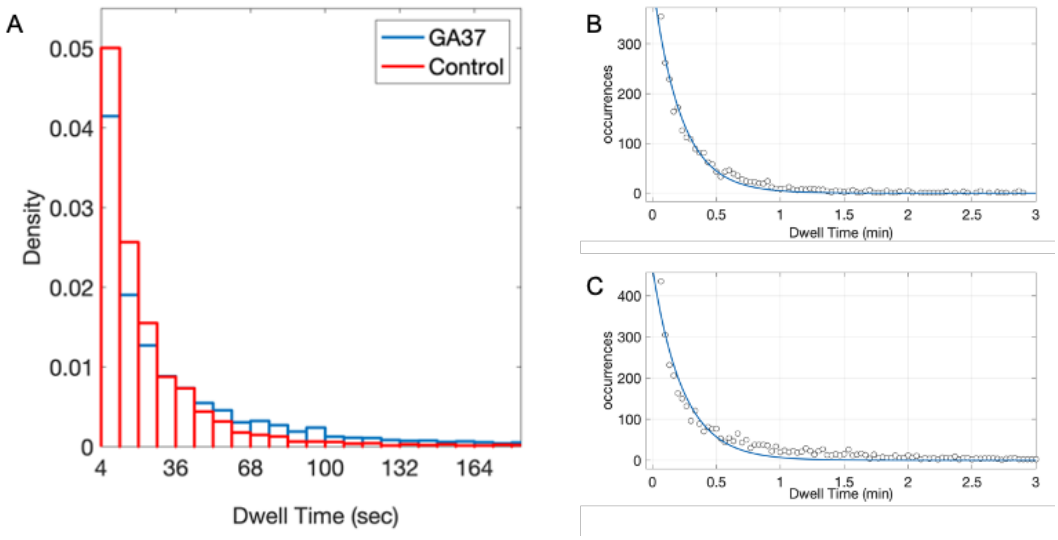


Figure 3.8.3. dwell time distribution and curve fitting over the distribution for GAr37 interacting with ClpX.
A) Normalized dwell time distribution histogram overlay of GAr37 (blue) and the control substrates interacting with ClpX. The control dataset was reused from the 48-a.a. tail substrate shown in Figure 3.5.2(D).
B) Fitting for ClpX-Control interaction (τ = 15.9 ± 0.2 sec; R-squared > 0.99)
C) Fitting for ClpX-GAr37 interaction (τ = 14.29 ± 1.1 sec; R-squared = 0.94)

3.9 Summary of results

A simple TIRF assay was established to specifically measure the release of substrate after its unfolding by ClpX was stalled due to structural stability. The system was tested for its passivation capacity and showed promising ability to measure the lifetime of individual ClpX-substrate complexes.

Using this TIRF system, in the absence of ClpP, when ClpX attempts to unfold a ssrA-tagged substrate, the k_{rel} can be reduced by extending the length of the C-terminal unstructured region, but the efficacy of such alterations plateaus. For an MTX-stabilized DHFR substrate with 48 a.a. tails, the mean lifetime of ClpX-substrate complex is about 15-20 seconds, as confirmed by both the TIRF assay as well as the SPR.

The initial binding of the ssrA-tagged substrate likely requires ATP hydrolysis. When substrate binding is impeded by a slowly hydrolysable ATPase homolog, substrate dwell time is significantly shortened. However, if ATP inhibition is at a level that allows substrate

binding, substrate dwell time can be prolonged by the inhibition of ATPase activity using ATPγS.

When ClpX is associated with ClpP, lengthening the substrate tail stops increasing dwell time when the tail is extended to 22 a.a. long. Compared to one-exponential decay, a two-exponential decay model is better suited to describe the dwell time distribution of substrates binding to ClpXP, but not enough parameters are known to restrict the model.

In terms of the effects of the amino acid sequence of the tail region on the k_{rel} of the substrate, it was confirmed at single molecule level that glycine-alanine repeats do not make the substrate more likely to be released. However, the bulk solution degradation assay cannot establish dramatic changes in the overall degradation rate, likely due to limitations of the substrate design.

In the next chapter, the implications of the findings, conclusions, and proposed hypotheses for future studies will be discussed.

Chapter 4
Discussions and Conclusions

4.1 The validity and limits of the TIRF imaging assay

The only parameter measured by the TIRF imaging assay utilized here is the dwell time of puncta appearing on a coverslip. To assess whether biochemical dwell times are accurately approximated by experimentally measured by punctum lifetime distributions, several conditions must be satisfied. Some of these conditions were directly verified, and others were made plausible based on known properties of ClpX, as discussed below.

The specificity of interaction between substrate and the ClpX (or ClpXP) in the TIRF assay used in this work was based on the differential binding of the ssrA and the ssrADD mutant to the coverslip surface. Unlike a FRET assay, wherein an interaction between two particles is confirmed by energy transfer between fluorophore pairs in close proximity, in our TIRF assay the identity of individual puncta and the causes that lead to its appearance and disappearance cannot be exactly determined for each individual case. For example, a punctum could appear due to non-specific substrate binding to the coverslip, perhaps due to insufficient passivation; a punctum could also be an optical impurity that can be mistaken for a fluorescently labeled substrate. This disadvantage was a tradeoff for the benefit of using multiple copies of fluorophore dyes for each substrate molecule to mitigate photobleaching and improve brightness of each punctum. This also simplifies the experimental design, which, for a question that had not been previously explored thoroughly at single molecule level, was an advantage. The TIRF assay certainly has room for improvement. I believe that the development of the current assay has laid a good foundation for future work.

One aspect of the TIRF assay development and validation concerns ways to address the various uncertainties caused by the limitations of the current design. The primary source of uncertainty, the nonspecific binding of particles to the coverslip surface, needs to be controlled by experimental methods. This was achieved by including a ssrADD mutant, which does not bind to ClpX. Substrates bearing the inactive mutant ssrADD degron provide an excellent negative control. They can thus be used to map out the areas of the coverslip that are not fully passivated. However, in a single molecule test, it is impossible to fully eliminate non-specific binding; as shown in Figures 3.2.4 and 3.2.5, the current passivation method typically had a background density of ranging from 10 to 30 puncta per frame. Because the process of finding a clean area for imaging was done manually, this process cannot be described as quantitative or unbiased. The criteria for finding a

good field of view for imaging based on the low density of puncta defined by ssrADD mutant binding were that the background dots were dispersed instead of aggregated, that the punctum density should be below 20 per frame, and that no rapid turnover of new ssrADD puncta were observed within a brief period. All of these criteria were assessed during imaging experiments by eye, and it could lead to inconsistencies.

In practice, coverslips were incubated with both the ssrA substrates and ssrADD mutant simultaneously. By labeling the two probes with different fluorophores, one could monitor specific bindings of the ssrA substrate after assessing the level of background non-specific binding in a field of view. To avoid confusing the data processing script, when switching the channel to the ssrA substrate for substrate imaging, it was desirable to not have a highly populated field of view, which complicates tracking of contiguous puncta which establishes the parameter of puncta lifetimes. Therefore, the punctum density reported for the ssrA substrate, which is usually around 100 puncta per frame, was selected during each imaging experiment, and should not be counted as a quantitative parameter for substrate binding capability. However, while it was difficult to use punctum density quantitatively, it was still apparent when a substrate demonstrated difficulty in binding to the coverslip, such as when punctum density would drop below the criteria set for acceptable background density of the ssrADD mutant. This happened for the imaging experiment with 2mM ATPγS as the sole nucleotide source, as shown in Section 3.6, and was one of the reasons that this nucleotide condition was not used further.

A second source of uncertainty is that the specific cause for the termination of an individual observed event is not clear. Causes termination could be due to biological or technical reasons. For single molecule microscopy, photobleaching of fluorophores is the primary source of artifact can lead to loss of puncta. But in practice, photobleaching was less of a concern after precautionary measures were taken. As shown in Figure 3.5.5, twice the amount of exposure did not shorten overall measured dwell time statistically. Thus, the uncertainty can be narrowed down to the exact mechanism that could cause the substrate to release.

To illustrate this point, consider the problem of how the substrate could be released, where “release” is defined broadly as the complete dissociation of the substrate from ClpX. It can be argued that based on the geometry of the central pore, which is narrowest at its opening, the probability for a bound substrate backing out from the central pore could be different than the probability for a bound substrate to exit from the side of hexamer, if a dynamic instability of the hexamer could allow a large “seam” to be opened. It is probable that many such alternative hypotheses can be proposed, but in this thesis, the release was described using a single rate constant k_{rel}, with the key assumption that the substrate must be pulled tight against ClpX and cannot be unfolded. As shown in Section 3.1, if

there are any alternative means of substrate release, k_{rel} would be a function of several rate constants; but the more potential pathways of release are included in the model, the more restraining functions are required to fit the population distribution which would reduce the sensitivity of the assay to parse out any of these parameters, unless these different release pathways can be identified and measured directly. Unfortunately, the TIRF assay at the current stage, which only measures the dwell time of the substrate, is not equipped to explore such possibilities.

Nonetheless, the thesis has demonstrated that even modeling the entire process using a single rate constant k_{rel}, many informative observations can be obtained that will direct future investigations.

4.2 Loss of grip by tyrosine loops, and how it can be compensated by longer tail

There were two notable single molecule studies on ClpXP slippage, conducted by Maillard et al. (2011) and Iosefson et al. (2015). Similar techniques used in these two works; in this section the method is introduced based on Iosefson et al. (2015). The results of Maillard et al. will be covered in Section 4.4, where the role of ClpP binding is discussed.

In the study of Iosefson et al. optical tweezers were used to show that ClpX would slip along the substrate during substrate translocation and unfolding. The substrate was pulled by ClpX in one direction, and by optical tweezers in the opposite direction at a force below stalling force. In this setting, ClpX is still able to translocate the substrate into its central pore even with the counterforce applied by the optical traps. Occasionally, the substrate was pulled back in the direction opposite to translocation by the counterforce applied by the optical tweezers. This sudden, backward movement was interpreted as ClpX losing grip on the substrate, which was referred to as a slippage event by the authors. In this experiment, with counterforce at 10 pN on average, ClpX, with all six tyrosine loops intact, slipped on average 5 times per minute during unfolding, but only 2 events per minute during translocation. Slippage distance—defined as the length of the substrate that was pulled out from ClpX before ClpX could regain its grip—was also larger during pre-unfolding dwell (close to 80 a.a) than during translocation (around 50 a.a.). When mutating one or more tyrosine loops by Y153A mutation, slippage events became more frequent during unfolding but not during translocation.

The above result from Iosefson et al. implied that slippage is determined by two factors. First, it was associated with how much resistance was met by the tyrosine loops during their paddling motion (thus during the pre-unfolding stage, more slippage events were

observed). Second, it was associated with the interaction between the tyrosine loops and the substrate, with more Y153A mutations reducing the interaction.

However, the technical limitations of optical traps systems should be noted. Here, the specific definition of slippage in this study was a combination of ClpX tyrosine loops losing grip over the substrate, and the optical tweezers pulling the substrate; it was difficult to predict how slippage distance would change if the optical tweezers counterforce had not been present, which would be more relevant to physiological conditions.

Importantly, slippage distance is a parameter determining whether ClpX can regain grip before a substrate fully slips out. In the absence of a counterforce from the optical tweezers, the slippage distance measured in a.a. might be changed in unpredictable ways. This is due to two factors. The absence of counterforce would reduce the force vector in the direction of releasing the substrate, thus potentially reducing the slippage distance. However, the slippage distance was reported as the amino acid distance which was converted from a physical distance measured in nm; when a peptide is not stretched by a counterforce, the number of residues per nm is increased, thus requiring a longer tail to span the same physical distance, making the slippage distance potentially longer according to an amino acid metric. Therefore, the optical tweezers data alone cannot fully predict the probability of the substrate release from each slippage event under physiological conditions.

The TIRF experiment shown in this thesis can fill some of the gaps from optical tweezers studies. To illustrate the point, the definitions used by Iosefson et al. can be borrowed to describe a model for the ClpX-substrate interaction of our TIRF experiment. Two important definitions are the slippage frequency (using Iosefson et al.'s definition), and slippage distance. As shown in Figure 4.2.1, when ClpX stalls at the folded domain, slippage can occur at a certain frequency, with the slippage distance defined as the distance the substrate can slip assuming that the tail region is infinitely long.

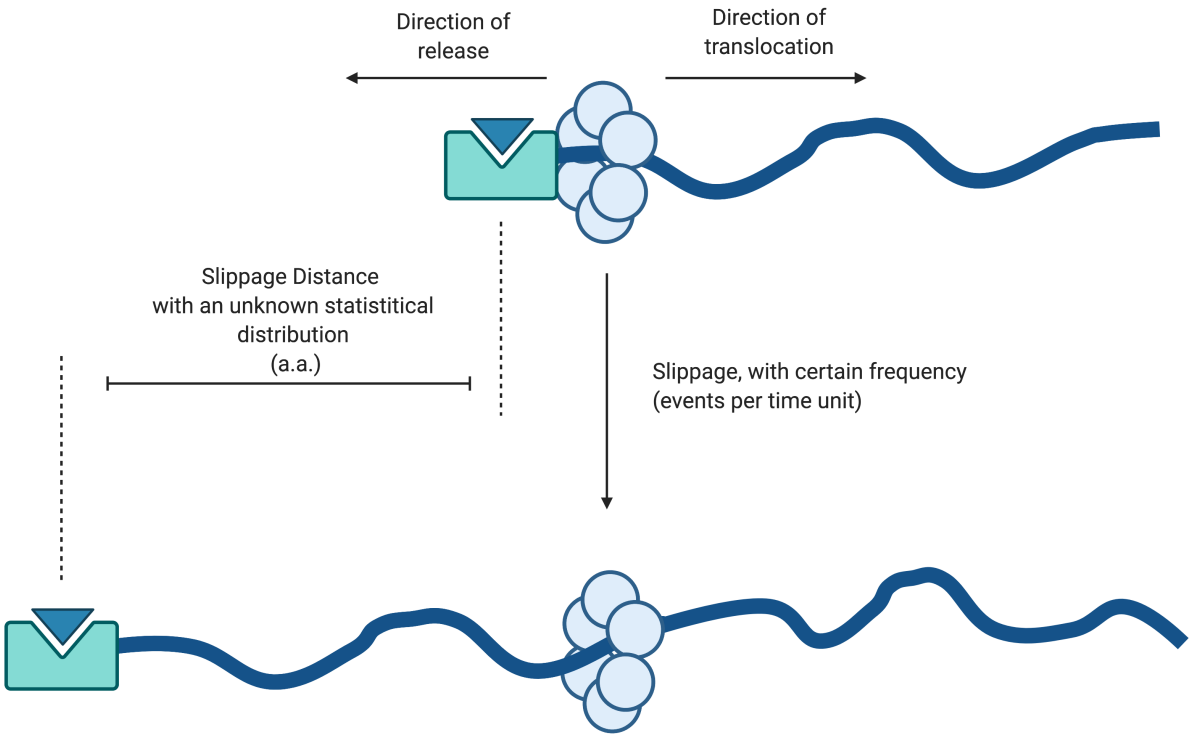


Figure 4.2.1 Slippage model for the TIRF experiment, defined using parameters introduced by Iosefson et al. (2015).

In this naive model, slippage frequency is related to the frequency that interaction between the tyrosine loops and the substrate is broken, while slippage distance is dependent on how fast ClpX can regain grip and how far the substrate can physically diffuse in the release direction when not gripped by ClpX.

In Section 3.5, I showed that progressively extending the tail of the substrate from 11 a.a. to 85 a.a. extends the substrate dwell time (from 3 sec to 17 sec) when it binds to ClpX, although the dwell time is not prolonged as much once the tail length exceeds 48 a.a. In Section 1.5, I propose that loss of interaction between the tyrosine loops and substrate does not necessarily always lead to the dissociation of the substrate. When explained by the model shown in Figure 4.2.1, it became clear that extending the tail beyond the slippage distance will promote ClpX re-establishing interaction with the substrate after a slippage. When the slippage frequency is constant, and the tail of the substrate is much shorter than the slippage distance, the substrate is more likely to be released.

To re-state my hypothesis proposed in Section 1.5 using the language from the above model, the slippage frequency is higher than the rate of substrate dissociation, because ClpX can regain grip after a slippage event. This means that if slippage is solely measured

by complete dissociation of a degradation intermediate, this measurement likely underestimates the true rate of slippage.

In the cryo-EM structure of ClpXP by Ripstein et al. (2020), ClpXP was stalled in the pre-unfolding state with the substrate—GFP-ssrA—bound in the central pore. The structure showed that the 11 a.a. ssrA tag was sufficient to span the entire length of the spiral staircase. This distance was also confirmed by structures from Fei et al. (2020), although in this case the identity of the substrate—presumed to be an endogenous remnant—was unknown.

Based on structural data and the model described in Figure 4.2.1, if the substrate with an 11 a.a.-tail slips from ClpX, ClpX is not likely to recover its grip. Therefore, the dwell time obtained from 11 a.a. tail substrate should reflect the basal probability of ClpX losing grip in the absence of ClpP. If ClpX rarely slipped, the 11 a.a. tail would be sufficient to support a much longer interaction. Instead, the mean lifetime of the ClpX-substrate complex measured by TIRF is about 3 sec, much shorter than the expected lifetime derived from the slippage frequency reported by Iosefson et al. (about 12 sec).

Interestingly, when the tail length is progressively increased to 33 a.a., the mean lifetime (around 13 sec) approaches what Iosefson et al. reported in their studies. The fast increase in average dwell time by extending the tail from 11 a.a. to 33 a.a. indicates that there are large variations in slippage distances; if slippage distance were fixed around 80 a.a. with small variations, extending the tail this way would not have much effect on substrate dwell time until tail length reaches beyond this value. This implies that the optical tweezers measurements from Iosefson et al. perhaps only recorded the more prominent slippage events.

When the tail is further extended, the dwell time plateaus at about 15-20 seconds. It is interesting that a very long tail in the substrate fails to dramatically improve substrate retention in the TIRF experiment. My hypothesis is that instead of having the slippage distance peaked around a single point, as reported by Iosefson et al. (2015), the slippage distance has a broad distribution with multiple peaks, resulting from the mechanisms by which the slippages occur. For example, there could be a class of slippage events, for which the slippage distance could be measured above 85 amino acids, such that even the longest tail length tested would not be sufficient to recover grip. While this hypothesis is not supported by results from Iosefson et al., it is partially supported by results from Maillard et al. (2011). In their study, Maillard et al. measured slippage distance for ClpX stalled at pre-unfolding dwell when attempting to unfold GFP. Their measurements

showed that slippage distance had a high variance, but mainly peaked at 30-50 a.a., with some slippage events well above 100 a.a., some as long as 300 a.a.

The different results from the two studies mentioned above reflect that the statistical model describing the distributions of the slippage distance is not well characterized. Unfortunately, the data presented in the thesis do not offer enough mechanistic insight to further explain the source of variations for the distribution of slippage distances. A hypothetical cause for a large-scale slippage could be that when ClpX tries to unfold a very stable substrate, the steric clash between the substrate and ClpX could sometimes destabilize the structure of ClpX instead of the substrate, to an extent that ClpX re-stabilization and recovery of grip would be impossible. However, without a method to monitor ClpX dynamics in real time during the substrate unfolding process, I find it challenging to test this hypothesis.

In summary, the TIRF result indicates that when the ClpX cannot regain grip after slippage, the substrate dissociates from ClpX more easily, indicating that basal slippage frequency is high; the substrate dwell time can be prolonged by extending the length of the tail, indicating that release is a function of slippage distance.

4.3 The force delivered from ClpX to the gripped substrate might weaken ClpX-substrate interaction, when the folded substrate cannot be translocated further

It needs to be emphasized that there could be many potential causes that could break the interaction between the tyrosine loops and the substrate, such as external factors like thermodynamic fluctuations, or internal factors like instabilities caused by the conformational changes. The effect of ATPγS inhibition described in Section 3.6 indicates that one major cause for breaking such interaction is in the mechanical cycle of ClpX.

As introduced in Section 1.3.2, the optical tweezers studies indicated that ATPγS binding pauses the ClpX ATPase cycle, making the inhibited subunit unable to perform the conformational changes that would otherwise result in a step in the translocation process (Sen et al., 2013; Rodriguez-Aliaga et al., 2016). It can be inferred that during protein unfolding, when force is delivered by a conformational change of ClpX that moves the tyrosine loop, impairing conformational changes by ATPγS should prevent force delivery and freeze the complex. The results in Section 3.6 indicates that slowing the ATPase hydrolysis cycle results in a longer average dwell time for the substrate.

Considering this TIRF experiment observation and the optical tweezers measurements from Iosefson et al. showing that slippage events were more frequent during pre-unfolding dwell stage, I draw the conclusion that at least a portion of slippage events are caused by the force delivered from ClpX. However, the exact mechanism by which this force causes slippage is not clear. One hypothesis is that during pre-unfolding dwell, ClpX attempts to unfold the substrate but fails to move it, causing the pulling force to break the interaction between the pore loops and the substrate. An alternative explanation is that the DHFR domain clashes with the ClpX hexamer ring, destabilizing its structure, and causing a more "catastrophic" loss of grip. Once again, studying the structural dynamics of ClpX in real-time during substrate unfolding will provide the much-needed insight on this matter.

4.4 Role of ClpP on substrate slippage

It has been reported that ClpX alone does not unfold GFP as efficiently as ClpXP (Kim et al., 2000), implying the possibility that ClpP might change the dynamics of ClpX to allow improved substrate unfolding. At the single molecule level, Maillard et al. attempted to test this hypothesis by measuring how the provision of ClpP could change the “grip” of ClpX on substrate. The methodology used by Maillard et al. to assess slippage was later used by Iosefson et al. (2015), as was described in Section 4.2. One notable difference between the two studies was that Maillard et al. were able to apply less counterforce and were able to measure slippage of substrate from ClpP. Comparing ClpX and ClpXP, Maillard et al. argued that ClpXP was less prone to “slippage” compared to ClpX, assessed by the average time a slip happens after ClpX or ClpXP enters the pre-unfolding dwell stage; by this measure, ClpX was 5 to 10 times more prone to slippage compared to ClpP (Maillard et al., 2011).

In all optical tweezers measurements, the substrate by definition is a “long-tail” substrate. In Section 3.6 it was shown that for the 11 a.a.-tail substrate, its dwell time with ClpX could also be prolonged in the presence of ClpP (increased from 3.1 sec to 6 sec). Because the tail of the substrate is too short to have any direct contacts with ClpP, the observation must reflect an effect caused by changes in ClpX dynamics upon ClpP binding.

As described in Section 3.7, when the substrate tail length is extended to 22 a.a and beyond, TIRF measurement showed that for ClpXP, 22 a.a- tail substrate had similarly long dwell time (14.3 sec) compared to substrates with much longer tails (between 13.5 to 17 sec). This result is quite different from those measured in the absence of ClpP as shown in Section 3.5 (τ = 9.8 sec for 22 a.a.-tail substrate; τ is increased progressively to

17.5 sec for substrates with longer tails). There are three hypotheses to explain the reduced tail-length dependence when ClpP is present. First, ClpX dynamics could be changed by ClpP binding, such that the average slippage distance would be reduced to the point where tail-length would become irrelevant when extended beyond 22 a.a. long. Second, the presence of ClpP could reduce slippage frequency per unit time (the model was shown in Figure 4.2.1). The second hypothesis can be further subdivided into two non-mutually exclusive scenarios. In the first scenario, slippage frequency could be reduced by changes in ClpX dynamics upon ClpP binding. In the second scenario, slippage frequency could be reduced when the tail of the substrate could reach ClpP and non-specifically interact with ClpP.

The present data cannot effectively distinguish between these two main hypotheses. To test the first hypothesis, because slippage distance is a concept originated from optical tweezers studies, it might be tempting to test the hypothesis with an optical traps system. However, there are several technical challenges that could limit the use of the technique. First, the experiment must avoid nonspecific interactions with ClpP by using a short-tail substrate, but at the same time needs to have a sufficiently long-tail substrate to prevent the substrate from being released and terminating the measurements. This is a compromise unlikely to be achieved by optical traps systems. Therefore, I propose to further characterize the slippage distance in the presence of ClpP using the current TIRF system. This could be achieved by using substrates of tail lengths at the intermediate range between 11 a.a. and 22 a.a. The challenge is that the temporal resolution of our TIRF system might not be sufficient to resolve the differences between these substrates. However, the resolution might b improved by reducing the ambient temperature during imaging or using ATPγS.

If future structural and dynamics studies provide evidence showing any potential interactions between the substrate tail and ClpP, this will give weight to the hypothesis proposing a reduction in slippage frequency. The collar at the gate of ClpP, which is constructed from the N-terminal residues of ClpP protomers, is the narrowest point in the ClpP barrel, and thus should be the prime candidate region for potential interactions with the substrate. Unfortunately, structural and dynamics data are still limited on this subject, and one cannot make a convincing argument for either hypothesis. The only piece of structural evidence showing the substrate making contacts with ClpP comes from Fei et al. (2020), which lacks the required resolution to confirm an interaction between substrate and ClpP. In the case of 26S proteasomes, Andres de la Peña et al. (2018) resolved a cryo-EM structure with the substrate reached below the gate of the core particle. At the opened gate of the core particle is a hydrophobic collar made from the N-terminal region of the alpha-subunits of the core particles. Once again, while the substrate tail is shown to make contact with the hydrophobic collar, the resolution of the substrate is still very

poor and an interaction between the substrate and the hydrophobic collar is not confirmed. In the future, a cross-linking experiment can be used to examine this process, but even so, a confirmation of contact is not a good assessment for how much grip such a contact could produce.

In summary, results from this thesis show that ClpP binding to ClpX prolongs substrate dwell time and makes it independent of tail length when the length is greater than 22 a.a. long. However, the mechanism of the effect is unclear.

4.5 The inhibitory effect of glycine-alanine repeats (GAr) on protein degradation

It was debated between earlier works by Too et al. (2013) and Kraut et al. (2012) whether GAr sequence could impair substrate degradation by increasing the release of the substrate. Kraut (2013) proposed a revised analysis of the data of Too et al. (2013), which the authors accepted. The revised conclusion showed that the GAr sequence likely did not change the release rate of the substrate, although there is still disagreement on the extent of the effect and the underlying mechanism from Too el al. in subsequent discourses.

The discussions surrounding this question prompted us to assess the release rate of the DHFR-substrate with a GAr tail at the single molecule level. In Section 3.8, it was shown that the GAr15 sequence and GAr37 had little effect at changing substrate dwell time with ClpX compared to the control sequence. This result confirms the conclusions from Kraut et al. (2012; 2013) and Too et al. (2013).

However, conclusions from a recent study by Bell et al. (2019) on the effect of tail sequence for ClpX encouraged us to revisit the hypothesis on how a GAr sequence impairs degradation. As mentioned in Section 3.8, Bell et al. showed that while a tail made of glycine significantly reduces overall degradation rate of GFP, inserting multiple alanine residues into the glycine tail provides sufficient grip for degradation of GFP.

It should be discussed how the GAr sequence could have a differential effect on the degradation of GFP and DHFR. One hypothesis is that the GAr sequence requires the collaboration of a tightly folded domain to impede degradation. GFP is a relatively hard-to-unfold substrate (as practically measured by its rate of degradation by ClpXP, as

characterized by Nager et al., 2011), but when the substrate is more difficult to unfold than GFP (such as DHFR, as characterized by Lee et al., 2001), alanine might become insufficient to enhance degradation compared to bulkier amino acids.

To elaborate this point, we can consider a hypothetical substrate that has a "Goldilocks" stability for GAr sequences. For this substrate, it can be degraded by having a random diverse sequence in its tail, but cannot be unfolded when having a GAr tail. When using this "Goldilocks" substrate to test the effect of GAr in a degradation assay, the conclusion would be that the GAr sequence could significantly impair substrate degradation.

Now we consider the case when the stability of the hypothetical substrate increases further, such that even a random diverse tail sequence would not let the substrate be unfolded. When repeating the degradation assay with this new substrate, the conclusion would be that the GAr substrate has no effect on substrate degradation.

In the experiment done by Kraut et al. (2012), the *E. coli* DHFR domain unfolding by yeast proteasomes was slowed 15-fold when a glycine-rich-region (GRR) replaced a control sequence in the tail region. However, upon DHFR stabilization by NADPH, the GRR sequence only reduced degradation rate by about 31% for yeast proteasome, a number that is much closer to my measurements with ClpXP for the DHFR substrate with GAr15 tail (My measurement showed a 12% reduction; Section 3.8). In all the experiments referenced above (and including results from this thesis), the full spectrum of substrate stability has not been used to test the effect of amino acid sequence on substrate unfolding. This could lead to wrong generalizations.

A second problem to be discussed is the spacing of the GAr sequence from the folded domain. As mentioned in Section 3.8, the design of the GAr15 substrate was based on the previous works from Too et al. (2013). It was shown that degradation of the DHFR domain was most heavily inhibited when the GAr sequence was placed 22 a.a. from the folded domain. The spacing was rationalized as the distance between the site where ClpX paddled the substrate, and the location where unfolding was prevented. In other words, when the ClpX was stalled at the folded domain, ClpX should grab the area 22 a.a. away from the folded domain, as shown in the diagram below (Figure 4.5.1).

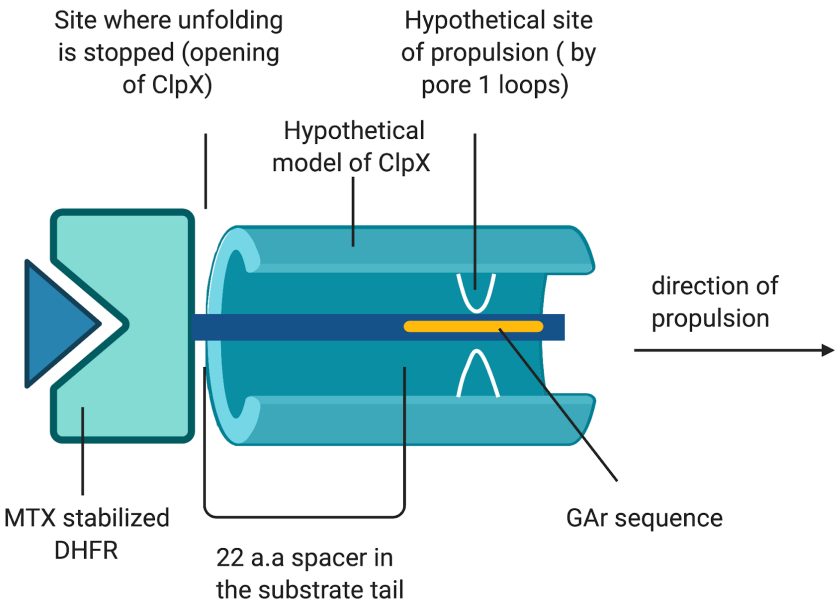


Figure 4.5.1: the hypothetical model from Too et al (2013) rationalizing how a 22 a.a. spacer between the DHFR domain and the GAr sequence promotes formation of degradation intermediates. The model proposes that the GAr sequence is imposed on the tyrosine loops when the substrate is held against ClpX. This model was first described for proteasomes same group (Hoyt et al., 2006).

Too et al. hypothesized that this placement of the GAr region would reduce the grip by the pore loops, thus impairing ClpX's ability to unfold the substrate. However, from cryo-EM structures by Ripstein et al. (2020), ClpX grips the substrate near the opening of the ClpX central pore. The site of propulsion is also functionally confirmed by results from Bell et al. (2019), which is mapped to be at 3 to 5 a.a. away from the folded domain. The rationalization by Too et al. to explain the 22 a.a. spacing is therefore not supported by recent studies. This raises a new question—what physical spacing within the ClpX-substrate complex determines the biochemical 22 a.a. optimum for spacing between GAr and folded domain?

Based on the cryo-EM structures, the 22 a.a. spacer would put the 10 a.a. GAr region below the ClpX hexamer ring, near the opening of ClpP, when translocation is stopped by the DHFR domain. This is supported by studies from Martin et al. (2008a), which showed that ClpXP with ATPase mutations releases folded substrates with their tails trimmed to 38 a.a.

If 38 a.a. is the distance from the opening of ClpX central pore to the proteolytic site, the 10 a.a. GAr should be closer to the ClpP proteolytic site than to the tyrosine loops. This

has led me to consider the possibility that the GAr sequence could interfere with the activity of ClpP rather than the grip between the substrate and ClpX.

One candidate target that could be interfered by the GAr sequence is the N-terminal region of ClpP that makes up the gate of the ClpP barrel. The gate region of ClpP has been shown to be a highly dynamic structure. In a study on ClpP gate allostery by Siavash Vahidi et al. (2018), the flexibility and heterogeneity of the collar elements were documented by NMR. Structures from Ripstein et al. (2020) showed that upon IGF loops binding on ClpP, the N-terminal region of the ClpP rigidifies into a β-hairpin, but the degree of rigidity is highly heterogeneous. The formation of the collar upon ClpX binding was also confirmed by structures from Fei et al. (2020) and Gatsogiannis et al. (2019). Because of the highly dynamic nature of its arrangement, I think it is plausible that non-specific interaction between the hydrophobic GAr sequence and the collar could occur. While transient interaction between GAr and the collar might not have any noticeable effect on ClpP activity, in the case of DHFR substrate the interaction is prolonged as a result of stalled ClpX, which could result in unexpected changes in ClpP dynamics.

In Section 3.8, it was shown that after assuring that ClpX interact with the GAr sequence by replacing the entire tail with GAr (in the GAr37 substrate), slippage was not increased by GAr37 in the absence of ClpP. This result strongly supports that glycine-alanine repeats do not lead to increased release rate for the substrate. However, this cannot be generalized for other low complexity sequences, as their effects have not yet been assessed using the same TIRF assay. In a recent single molecule FRET study on substrate interactions with proteasome, Jared Bard et al. (2019) tested a serine-rich (SR) tail sequence, and found that the insertion of the substrate into the proteasome central pore is significantly slowed down. Considering that the SR sequence has similar complexity to GAr or a PolyQ sequence, Kraut et al., Bell et al. and Jared Bard et al. showed that different low complexity sequences could have markedly different behaviors, and each inhibiting degradation by different means. The reduced speed of tail insertion for SR sequence implies that the initial translocation step for substrate engagement is impaired, which strongly suggests a change in the interaction between the tyrosine loops and the substrate. It would be informative to test how the SR sequence might change substrate dwell time when measured using the TIRF system described in this thesis, although cautions are needed to ensure that the initial translocation of the substrate before unfolding is not rate limiting in the overall dwell time.

4.6. Hypothetical Models

My thesis originated from our interest in studying how glycine-alanine repeats impair substrate degradation. The biochemical studies by Kraut et al. (2012) and Too et al. (2013) have shown that "slippery sequences" such as glycine-alanine repeats reduce the forward substrate degradation rates but do not significantly change substrate release rate when processed by AAA+ proteases. However, this outcome cannot be satisfactorily explained by existing mechanistic understandings. Based on the current model, ClpX unfolds the substrate by gripping the unstructured region and pulling it down into the central pore; in contrast, the release of the substrate is dependent on ClpX losing its grip on the substrate, so that the substrate can back out from the central pore. In trying to fit this model with the results from Kraut et al. and Too et al., glycine-alanine sequence makes the substrate difficult to grip by ClpX for unfolding, but at the same time, ClpX paradoxically grips the substrate effectively to prevent accidental release. Clearly, there was missing information in this model.

Recently a number of cryo-EM structures on AAA+ proteases, including ClpXP, became available. The majority of these studies propose a "hand-over-hand" model for the mechanical cycle of AAA+ ATPases (some highlights include Puchades et al., 2017; de la Peña et al., 2018; Ripstein et al., 2020). In this model, the six tyrosine loops in the central pore are arranged like a spiral staircase, where the bottom pore loop can move to the top of the staircase, driving the other five pore loops to take one step down the staircase. The substrate gripped by these downward moving pore loops is thus pulled down into the central pore (Figure 4.6.1; the same figure is previously shown in Chapter 1). Some higher resolution cryo-EM structures including those by de la Peña et al. and Fei et al. showed specific pairings of the tyrosine loops with the backbones and side chains of the substrate. These interactions were used to describe the nature of the grip that ClpX exerts on substrates.

A limitation shared by all these models is that they describe only the translocation process, and the grip between tyrosine loops and substrate is described as a static interaction. As thoroughly discussed by Maillard et al. in their investigation of ClpXP biophysical properties, the amino acid chain has a very short persistence length relative to the measured step size of the unfoldase, making it fundamentally different from other motor proteins that walk on stiff tracks like microtubules and DNA (Maillard et al., 2011). The irregularity and flexibility of the substrate makes a model derived from static structures less applicable; where most structural studies predicted a step size of about 2 residues per conformational change for ClpXP, optical tweezers studies discovered a much larger

step size when ClpXP was put in motion. Currently, this difference between structural studies and biophysical measurements has not been solved.

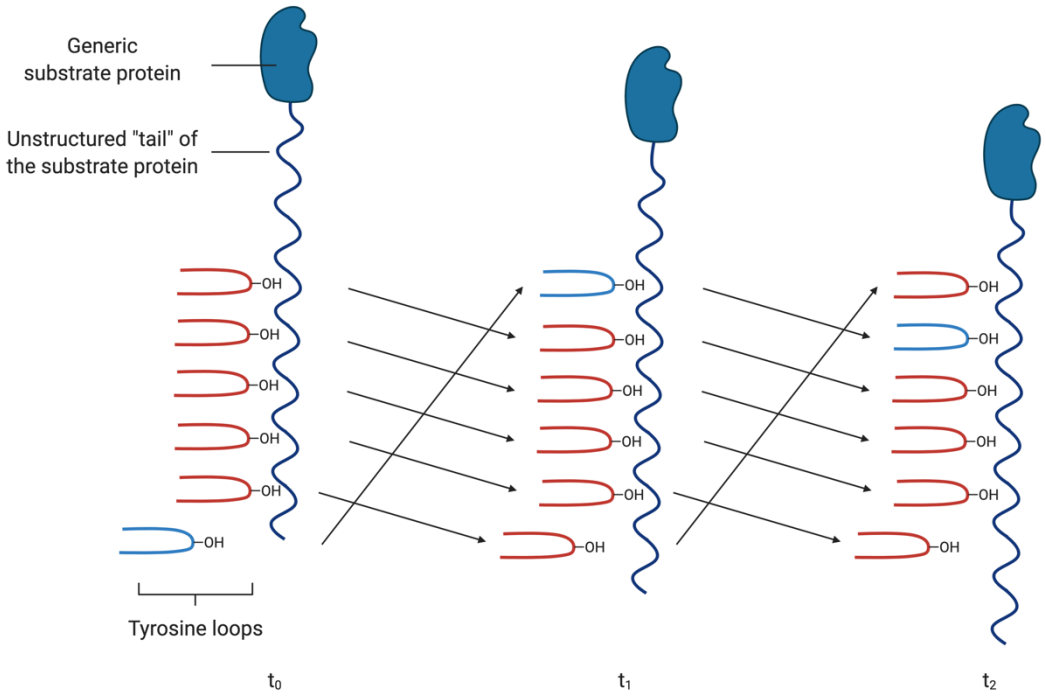


Figure 4.6.1 Hand over hand model for translocation.

Bell et al. (2019) highlighted this uncertainty of our current understandings of the process of substrate unfolding. One of their findings was that for a GFP substrate with a tail made of glycine residues, the substrate was unfolded most efficiently when the 4th position of the tail was replaced by a tyrosine, but unfolding was less efficient for a tyrosine replacement further away from the folded domain. Bell et al. proposed that the 4th-position residue interacts with a ClpX Pore 1 at the top of the spiral staircase, and the force delivered to the substrate was not equal among the Pore 1 loops. They proposed several models to explain why the upper Pore 1 loops were more efficient at delivering the unfolding force. In the first model, the upper pore loops could maintain a static grip with the substrate while the lower pore loops were more dynamic, potentially due to the increasing diameter of the central pore progressing from the top to the bottom of the spiral staircase. In the second model, the tail of the substrate acts like a spring, such that the

energy from a force delivered at the bottom of the spiral staircase would result in a longer expansion of the tail, which would absorb most of its energy.

My view of their discussion is that, aside from the slipperiness of the tail, the stability of the substrate, and the geometry of the ClpX central pore, two additional major factors could dictate efficiency and release of the substrate. The first factor is the elasticity or the pliability of the substrate; a highly elastic substrate can absorb the unfolding force from ClpX without getting unfolded. The second factor is the dynamics of ClpX during unfolding, which can be further divided into two smaller factors—the dynamics of the pore loops and the dynamics of the subunits. For pore loops, it is not clear whether they could maintain a static interaction with the gripped residue or not during unfolding; for ClpX subunits, it is not clear whether a substrate more rigid than ClpX could destabilize the ClpX hexameric ring before the ClpX could destabilize the substrate. As noted in Newton's third law, for every action, there is an equal and opposite reaction: ClpX and substrate act reciprocally on each other.

These uncertainties are not satisfactorily explained by models derived from the structural studies, partly because the "hand-over-hand" model proposed by structural studies is in essence a model of substrate translocation, which does not provide guidelines on systems where translocation is stalled by a physical barrier.

Based on the results presented in this thesis, I propose the following models that describe the unfolding process, with the focus on how ClpX could fail at unfolding under the hand-over-hand translocation model.

I propose that the grip between ClpX and substrate can be described as the static interaction between the tyrosine loops and the substrate tail region. The static interaction signifies that each tyrosine loop is engaged with a specific residue of the substrate, and this association remains fixed when the tyrosine loop moves down the spiral staircase. Assuming that ClpX uses the hand-over-hand mechanism for substrate translocation, during each duty cycle, the bottom tyrosine loop disengages from the substrate, moves to the top of the spiral staircase, and re-establishes a static interaction with a new substrate residue. This cycle will repeat until ClpX reaches the folded domain.

At this point, several outcomes could take place. The one that is easiest to consider is that the substrate is unfolded (leading to either one partially unfolded intermediate, or one fully unfolded product) as soon as ClpX pulls the substrate tail with the hand-over-hand

motion. When this occurs, ClpX continues to rapidly translocate the now unfolded (or partially unfolded) substrate into the central pore until reaching the next physical barrier.

It becomes much less certain what would happen when ClpX fails to unfold the substrate or an intermediate in the course of multiple ATPase cycles. Based on the TIRF measurements showing that slowing down ClpX ATPase cycle by ATPγS improves substrate dwell time (Section 3.5 and further discussed in Section 4.3), I propose that the force delivered from ClpX could destabilize the static interaction between the tyrosine loops and the substrate tail region, when the substrate cannot be moved further. However, there are still a couple of possibilities for how this destabilization could occur. I propose two scenarios, but these do not exclude other possibilities.

In the first scenario, when the substrate can be viewed as a rigid body, ClpX completes one mechanical cycle but fails to drive or deform the substrate (Figure 4.6.2). The net effect is that ClpX completes a cycle of conformational change, much as during unimpeded translocation, but ClpX does not change its relative position with the substrate. When this happens, it is necessary for several tyrosine loops to disengage their currently bound residues of the substrate and re-engage another set of residues in order to make room for the bottom tyrosine loop to move to the top. This model assumes that ClpX is capable of generating the force needed to break the interactions between tyrosine loops and the substrate. At the same time, this model requires such static interactions to be broken, in order for ClpX to complete a full mechanical cycle.

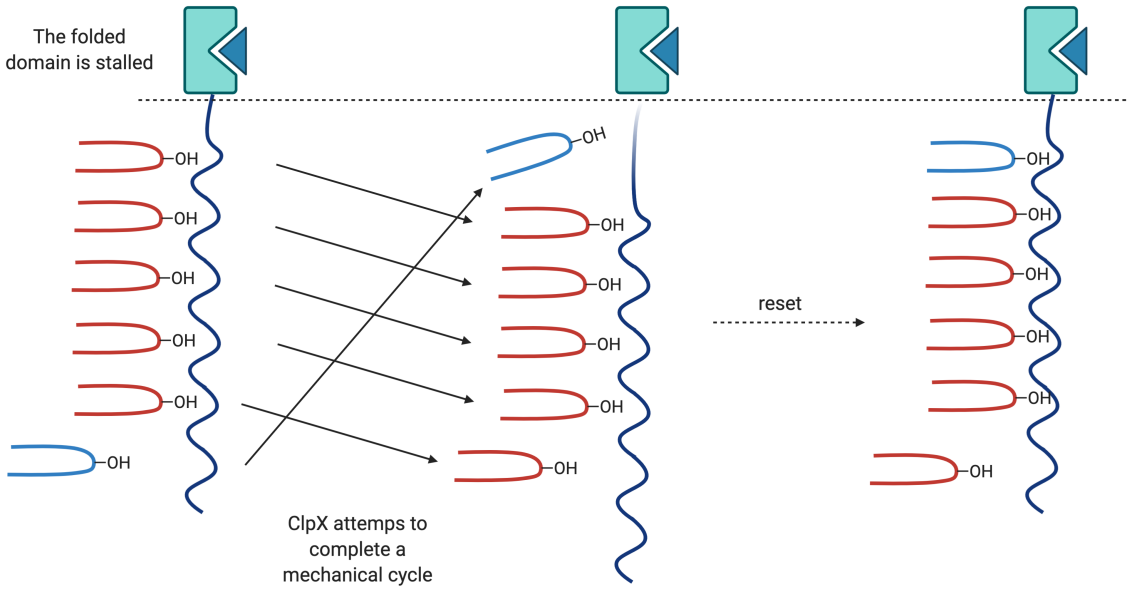


Figure 4.6.2: breaking and re-engaging static interactions between tyrosine pore loops and the substrate tail

In the second scenario, when a substrate is elastic, ClpX is able to complete a mechanical cycle without losing static interaction with the substrate, but still fails to unfold the substrate. (Figure 4.6.3). In this case, the substrate could continuously absorb energy like a stretched spring as ClpX continues to advance. At some point, the elasticity of the substrate would reach its limit, after which it could either be fully unfolded, or act as an obstacle as in the first scenario mentioned above; alternatively, it could release all the stored energy by springing back to its original shape, and the force released by this spring action could pull the substrate tail out of ClpX central pore, thus causing a slippage event. If the released energy from the spring action is large, this model predicts that the probability for ClpX to regain grip will be low.

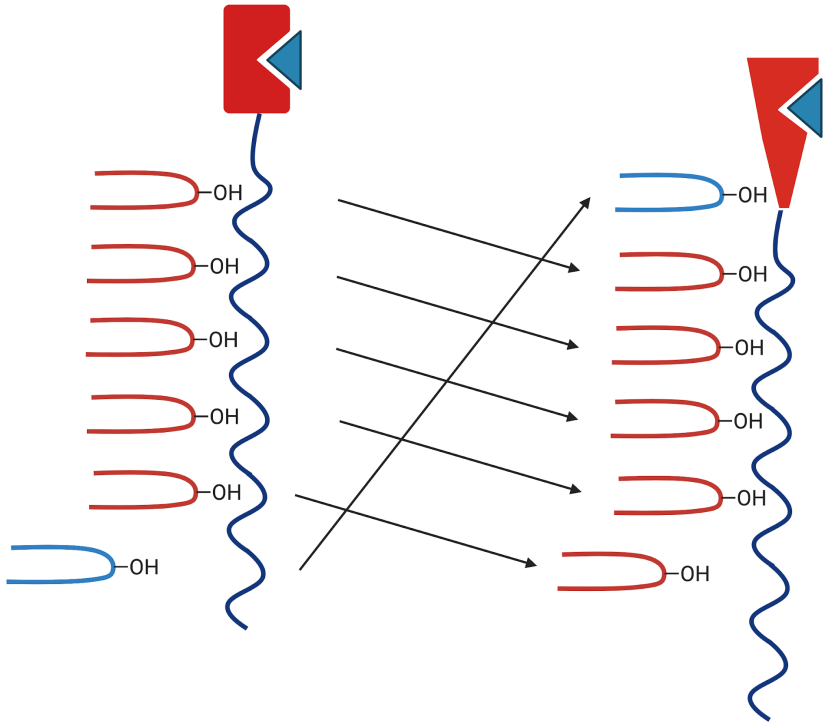


Figure 4.6.3: substrate is deformed but not fully unfolded. This could happen if ClpX completes a conformational change without breaking most of the static interactions between pore loops and substrates.

To help distinguish these two scenarios when ClpX stalls at a folded domain, one way would be to measure the force required to break the static interaction between ClpX and substrate, and compare it with the typical force ClpX can generate during an unfolding event. While force measurements have been extensively conducted using optical tweezers, the keyword here is "static", which has not been systematically studied. One conceptually simple experiment to do is to measure the force required to physically pull out the substrate when ClpX is paused by ATPγS. This has not been conventionally done

in optical tweezers experiments, for the obvious reason that pulling out the substrate could terminate each measurement, limiting the throughput of the assay. One technology suitable for this measurement is the centrifugal microscope developed by Wesley Wong et al. (Halvorsen and Wong, 2010), where the counterforce for pulling the engaged substrate out can be provided by centrifugation, which allows a much higher throughput to measure this simple biophysical parameter that has been overlooked before. Ideally, the slippage force for ATPγS-inhibited ClpX could be systematically measured for different amino acid sequences.

The effect of elastic substrate is much harder to test, because substrate dynamics would add much complexity to analysis of the system. One potential way to circumvent this problem is to measure the release rate of a stable substrate (like DHFR) with a spring-like alpha helix attached. The model proposed above predicts that a spring storing more energy should lead to more catastrophic slippage events when energy is released. However, while the spring-like properties of alpha-helices have been measured by biophysical methods like AFM (For example, by Wolny et al., 2014), how ClpX processes helical sequences is not well studied, making this experiment less appealing.

I emphasize again that the above two scenarios do not cover all possible situations when ClpX fails to unfold the substrate. One likely outcome is that ClpX might complete a futile ATPase cycle while failing to complete the mechanical cycle. Futile ATPase cycles are possible, because conformational locking studies have shown that the ATPase activity can be decoupled from the mechanical cycle by covalently locking ATPase motion (Stinson et al., 2013; also discussed in Section 1.3.1). Hypothetically, in a futile cycle ClpX must retain its static interaction with the substrate throughout; otherwise, motions of the tyrosine loops would not be restrained, and ClpX would complete the conformational change as described in the scenarios above. To test the presence of futile cycles, it is best to design a way to monitor conformational change during substrate unfolding. With the recently published cryo-EM structures, a FRET probe could be feasibly designed to monitor the hand-over-hand conformational change in real-time. A result showing a lack of conformational change during the pre-unfolding dwell stage would support this model.

The models proposed above focus on how static interaction between pore loops and substrates could be broken. However, it cannot effectively predict how the release of substrate would be affected after interaction was broken. The reason is that there could be multiple interactions that prevent the release of a substrate. For example, a longer tail in the substrate, or non-specific interaction between substrate and ClpP, could afford ClpX more time to re-establish static interaction between pore loops and substrate. Alternatively, a hydrophobic sequence might allow the tyrosine loops to glide over the

gripped area but still maintain the binding, as a result of hydrophobic interactions. Consequently, without knowing all the ways that ClpX can prevent substrate release after a loss of static interaction, it is difficult to predict how efficiently a substrate can be degraded.

The many branching possibilities in my model highlight the fact that many aspects of the unfolding process have not been fully explored. The combinations of substrate stabilities and tail sequences could be so numerous that typical biochemical studies cannot afford to explore all these factors at once. To narrow the number of plausible hypothetical models, I think it is best to characterize the biophysical process of substrate unfolding and fill the gap that the structural data cannot explain. This includes studying the dynamics of ClpX conformational change during unfolding, the static interaction between the pore loops and the substrate.

4.7 Concluding Remarks

In this thesis, the release of the substrates from ClpX and ClpXP monitored at single molecule level using TIRF microscopy. This method was established to specifically measure the release rate constant, or k_{rel}, of a non-degradable substrate, the MTX-stabilized DHFR, when it is stalled at the ClpX. The tail region of the DHFR substrate, where the tyrosine loops of ClpX grip during substrate unfolding, could be changed to perturb the interaction. Here, grip is defined as the static interaction between tyrosine loops and the protein substrate.

Four factors that could potentially influence the release rate of the substrate were explored. k_{rel} was shown to be reduced by extending the tail of the substrate, or by slowing down the ATPase cycle of ClpX using ATPγS, or by ClpX association with ClpP. However, k_{rel} was not reduced by changing the tail sequence to a glycine-alanine repeat. The results indicate that when a substrate is stalled at ClpX (such as when ClpX is working to unfold a challenging substrate), the frequency of losing grip by ClpX is correlated with the frequency of its ATPase cycles. This can be rationalized by a discord between the movement of the tyrosine loops and that of the substrate. When substrate stalls at the central pore, the freedom of the substrate to move with the tyrosine loops is significantly restricted. Forcibly moving the tyrosine loops after ATP hydrolysis thus could potentially break the interaction between the tyrosine loops and the substrate, leading to a slippage event.

When ClpX loses grip, the substrate could freely back out from the central pore. If the tail of the substrate is long, ClpX could regain grip before the substrate fully exits from the

central pore, thus preventing the release of the substrate after a slippage event. The probability that a substrate can be retained is related to the distribution of slippage distance, which is defined as the length of the substrate that slips out from ClpX central pore before ClpX regains its grip. The exact distribution of slippage distances still needs to be further characterized, but in principle a substrate with a tail length greater than the average slippage distance should be less likely to be released after a slippage event. On the flip side, ClpX is more likely to release a substrate with a shorter tail after a slippage event. ClpX losing its grip thus could be a common side-effect of the substrate unfolding process, but the release of the substrate is largely prevented when the substrate tail is long.

The glycine-alanine repeats (GAr), which are often associated with interruption of protein degradations, were shown to not increase the rate of substrate dissociation. This result suggests that GAr could disrupt protein degradation by directly inhibiting the activity of ClpP. However, the mechanisms by which ClpP regulates ClpX dynamics are not well understood. Here, association of ClpP was shown to reduce substrate release rate. This was likely achieved simultaneously by a reduced slippage frequency and a reduced slippage distance. Importantly, the kinetics of the substrate dissociation is not fully explained by a single rate constant. Further studies on ClpX dynamics upon ClpP association are required to interpret these observations.

REFERENCES

Aitken, C.E., Marshall, R.A., and Puglisi, J.D. (2008). An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments. Biophysical Journal *94*, 1826–1835.

Amor, A.J., Schmitz, K.R., Sello, J.K., Baker, T.A., and Sauer, R.T. (2016). Highly Dynamic Interactions Maintain Kinetic Stability of the ClpXP Protease During the ATP-Fueled Mechanical Cycle. ACS Chem. Biol. *11*, 1552–1560.

Amor, A.J., Schmitz, K.R., Baker, T.A., and Sauer, R.T. (2019). Roles of the ClpX IGF loops in ClpP association, dissociation, and protein degradation. Protein Science *28*, 756–765.

Aubin-Tam, M.-E., Olivares, A.O., Sauer, R.T., Baker, T.A., and Lang, M.J. (2011). Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. Cell *145*, 257–267.

Bard, J.A.M., Bashore, C., Dong, K.C., and Martin, A. (2019). The 26S Proteasome Utilizes a Kinetic Gateway to Prioritize Substrate Degradation. Cell *177*, 286-298.e15.

Barkow, S.R., Levchenko, I., Baker, T.A., and Sauer, R.T. (2009). Polypeptide Translocation by the AAA+ ClpXP Protease Machine. Chemistry & Biology *16*, 605–612.

Bewley, M.C., Graziano, V., Griffin, K., and Flanagan, J.M. (2006). The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. Journal of Structural Biology *153*, 113–128.

Bewley, M.C., Graziano, V., Griffin, K., and Flanagan, J.M. (2009). Turned on for degradation: ATPase-independent degradation by ClpP. Journal of Structural Biology *165*, 118–125.

Burton, R.E., Baker, T.A., and Sauer, R.T. (2003). Energy-dependent degradation: Linkage between ClpX-catalyzed nucleotide hydrolysis and protein-substrate processing. Protein Sci. *12*, 893–902.

Cordova, J.C., Olivares, A.O., Shin, Y., Stinson, B.M., Calmat, S., Schmitz, K.R., Aubin-Tam, M.-E., Baker, T.A., Lang, M.J., and Sauer, R.T. (2014). Stochastic but Highly Coordinated Protein Unfolding and Translocation by the ClpXP Proteolytic Machine. Cell *158*, 647–658.

Fei, X., Bell, T.A., Jenni, S., Stinson, B.M., Baker, T.A., Harrison, S.C., and Sauer, R.T. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. ELife *9*, e52774.

Flynn, J.M., Levchenko, I., Seidel, M., Wickner, S.H., Sauer, R.T., and Baker, T.A. (2001). Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. Proceedings of the National Academy of Sciences *98*, 10584–10589.

Gatsogiannis, C., Balogh, D., Merino, F., Sieber, S.A., and Raunser, S. (2019). Cryo-EM structure of the ClpXP protein degradation machinery. Nat Struct Mol Biol *26*, 946–954.

Glynn, S.E., Martin, A., Nager, A.R., Baker, T.A., and Sauer, R.T. (2009). Structures of Asymmetric ClpX Hexamers Reveal Nucleotide-Dependent Motions in a AAA+ Protein-Unfolding Machine. Cell *139*, 744–756.

Glynn, S.E., Nager, A.R., Baker, T.A., and Sauer, R.T. (2012). Dynamic and static components power unfolding in topologically closed rings of a AAA+ proteolytic machine. Nat Struct Mol Biol *19*, 616–622.

Gottesman, S., Roche, E., Zhou, Y., and Sauer, R.T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes & Development *12*, 1338–1347.

Halvorsen, K., and Wong, W.P. (2010). Massively Parallel Single-Molecule Manipulation Using Centrifugal Force. Biophysical Journal *98*, L53–L55.

Hersch, G.L., Baker, T.A., and Sauer, R.T. (2004). SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. Proceedings of the National Academy of Sciences *101*, 12136–12141.

Hersch, G.L., Burton, R.E., Bolon, D.N., Baker, T.A., and Sauer, R.T. (2005). Asymmetric Interactions of ATP with the AAA+ ClpX6 Unfoldase: Allosteric Control of a Protein Machine. Cell *121*, 1017–1027.

Hoyt, M.A., Zich, J., Takeuchi, J., Zhang, M., Govaerts, C., and Coffino, P. (2006). Glycine–alanine repeats impair proper substrate unfolding by the proteasome. EMBO J *25*, 1720–1729.

Hua, B., Han, K.Y., Zhou, R., Kim, H., Shi, X., Abeysirigunawardena, S.C., Jain, A., Singh, D., Aggarwal, V., Woodson, S.A., et al. (2014). An improved surface passivation method for single-molecule studies. Nat Methods *11*, 1233–1236.

Hwang, W., and Lang, M.J. (2013). Nucleotide-Dependent Control of Internal Strains in Ring-Shaped AAA+ Motors. Cel. Mol. Bioeng. *6*, 65–73.

Iosefson, O., Olivares, A.O., Baker, T.A., and Sauer, R.T. (2015). Dissection of Axial-Pore Loop Function during Unfolding and Translocation by a AAA+ Proteolytic Machine. Cell Reports *12*, 1032–1041.

Johnson, D.S., Jaiswal, J.K., and Simon, S. (2012). Total Internal Reflection Fluorescence (TIRF) Microscopy Illuminator for Improved Imaging of Cell Surface Events. Current Protocols in Cytometry *61*.

Joshi, S.A., Hersch, G.L., Baker, T.A., and Sauer, R.T. (2004). Communication between ClpX and ClpP during substrate processing and degradation. Nat Struct Mol Biol *11*, 404–411.

Keiler, K.C., Waller, P.R.H., and Sauer, R.T. (1996). Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA. Science *271*, 990–993.

Kenniston, J.A., Baker, T.A., Fernandez, J.M., and Sauer, R.T. (2003). Linkage between ATP Consumption and Mechanical Unfolding during the Protein Processing Reactions of an AAA+ Degradation Machine. Cell *114*, 511–520.

Kenniston, J.A., Baker, T.A., and Sauer, R.T. (2005). Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. Proceedings of the National Academy of Sciences *102*, 1390–1395.

Kim, D.Y., and Kim, K.K. (2003). Crystal Structure of ClpX Molecular Chaperone from *Helicobacter pylori*. J. Biol. Chem. *278*, 50664–50670.

Kim, Y.-I., Burton, R.E., Burton, B.M., Sauer, R.T., and Baker, T.A. (2000). Dynamics of Substrate Denaturation and Translocation by the ClpXP Degradation Machine. Molecular Cell *5*, 639–648.

Kim, Y.-I., Levchenko, I., Fraczkowska, K., Woodruff, R.V., Sauer, R.T., and Baker, T.A. (2001). Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. Nature Structural Biology *8*, 4.

Kirstein, J., Hoffmann, A., Lilie, H., Schmidt, R., Rübsamen-Waigmann, H., Brötz-Oesterhelt, H., Mogk, A., and Turgay, K. (2009). The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. EMBO Mol Med *1*, 37–49.

Koide, S., and Sidhu, S.S. (2009). The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. ACS Chem. Biol. *4*, 325–334.

Kraut, D.A. (2013). Slippery Substrates Impair ATP-dependent Protease Function by Slowing Unfolding. J. Biol. Chem. *288*, 34729–34735.

Kraut, D.A., Israeli, E., Schrader, E.K., Patil, A., Nakai, K., Nanavati, D., Inobe, T., and Matouschek, A. (2012). Sequence- and Species-Dependence of Proteasomal Processivity. ACS Chem. Biol. *7*, 1444–1453.

Lee, B.-G., Park, E.Y., Lee, K.-E., Jeon, H., Sung, K.H., Paulsen, H., Rübsamen-Schaeff, H., Brötz-Oesterhelt, H., and Song, H.K. (2010). Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. Nat Struct Mol Biol *17*, 471–478.

Lee, C., Schwartz, M.P., Prakash, S., Iwakura, M., and Matouschek, A. (2001). ATP-Dependent Proteases Degrade Their Substrates by Processively Unraveling Them from the Degradation Signal. Molecular Cell *7*, 627–637.

Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald-Mullen, P.M., Klein, G., Kurilla, M.G., and Masucci, M.G. (1995). Inhibition of antigen processing by the internal repeat region of the Epstein–Barr virus nuclear antigen-1. Nature *375*, 685–688.

Maillard, R.A., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C.M., Hodges, C., Martin, A., and Bustamante, C. (2011). ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates. Cell *145*, 459–469.

Martin, A., Baker, T.A., and Sauer, R.T. (2005). Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. Nature *437*, 1115–1120.

Martin, A., Baker, T.A., and Sauer, R.T. (2008a). Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. Nat Struct Mol Biol *15*, 1147–1151.

Martin, A., Baker, T.A., and Sauer, R.T. (2008b). Diverse Pore Loops of the AAA+ ClpX Machine Mediate Unassisted and Adaptor-Dependent Recognition of ssrA-Tagged Substrates. Molecular Cell *29*, 441–450.

Nager, A.R., Baker, T.A., and Sauer, R.T. (2011). Stepwise Unfolding of a β Barrel Protein by the AAA+ ClpXP Protease. Journal of Molecular Biology *413*, 4–16.

Nassif, N.D., Cambray, S.E., and Kraut, D.A. (2014). Slipping up: Partial substrate degradation by ATP-dependent proteases: Partial Substrate Degradation by ATP-Dependent Proteases. IUBMB Life *66*, 309–317.

Olivares, A.O., Kotamarthi, H.C., Stein, B.J., Sauer, R.T., and Baker, T.A. (2017). Effect of directional pulling on mechanical protein degradation by ATP-dependent proteolytic machines. Proc Natl Acad Sci USA *114*, E6306–E6313.

Ortega, J., Singh, S.K., Ishikawa, T., Maurizi, M.R., and Steven, A.C. (2000). Visualization of Substrate Binding and Translocation by the ATP-Dependent Protease, ClpXP. Molecular Cell *6*, 1515–1521.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol *24*, 79–88.

de la Peña, A.H., Goodall, E.A., Gates, S.N., Lander, G.C., and Martin, A. (2018). Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis–driven translocation. Science *362*, eaav0725.

Pickart, C.M., and Cohen, R.E. (2004). Proteasomes and their kin: proteases in the machine age. Nat Rev Mol Cell Biol *5*, 177–187.

Puchades, C., Rampello, A.J., Shin, M., Giuliano, C.J., Wiseman, R.L., Glynn, S.E., and Lander, G.C. (2017). Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. Science *358*, eaao0464.

Ripstein, Z.A., Huang, R., Augustyniak, R., Kay, L.E., and Rubinstein, J.L. (2017). Structure of a AAA+ unfoldase in the process of unfolding substrate. ELife *6*, e25754.

Ripstein, Z.A., Vahidi, S., Houry, W.A., Rubinstein, J.L., and Kay, L.E. (2020). A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. ELife *9*, e52158.

Roche, E.D. (1999). SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. The EMBO Journal *18*, 4579–4589.

Rodriguez-Aliaga, P., Ramirez, L., Kim, F., Bustamante, C., and Martin, A. (2016). Substrate-translocating loops regulate mechanochemical coupling and power production in AAA+ protease ClpXP. Nat Struct Mol Biol *23*, 974–981.

Sen, M., Maillard, R.A., Nyquist, K., Rodriguez-Aliaga, P., Pressé, S., Martin, A., and Bustamante, C. (2013). The ClpXP Protease Unfolds Substrates Using a Constant Rate of Pulling but Different Gears. Cell *155*, 636–646.

Stinson, B.M., Nager, A.R., Glynn, S.E., Schmitz, K.R., Baker, T.A., and Sauer, R.T. (2013). Nucleotide Binding and Conformational Switching in the Hexameric Ring of a AAA+ Machine. Cell *153*, 628–639.

Szyk, A., and Maurizi, M.R. (2006). Crystal structure at 1.9Å of E. coli ClpP with a peptide covalently bound at the active site. Journal of Structural Biology *156*, 165–174.

Tian, L., Holmgren, R.A., and Matouschek, A. (2005). A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-κB. Nat Struct Mol Biol *12*, 1045–1053.

Too, P.H.-M., Erales, J., Simen, J.D., Marjanovic, A., and Coffino, P. (2013). Slippery Substrates Impair Function of a Bacterial Protease ATPase by Unbalancing Translocation *versus* Exit. J. Biol. Chem. *288*, 13243–13257.

Voges, D., Zwickl, P., and Baumeister, W. (1999). The 26S Proteasome: A Molecular Machine Designed for Controlled Proteolysis. Annu. Rev. Biochem. *68*, 1015–1068.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO Journal *1*, 945–951.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in Python. PeerJ *2*, e453.

Wang, J., Hartling, J.A., and Flanagan, J.M. (1997). The Structure of ClpP at 2.3 Å Resolution Suggests a Model for ATP-Dependent Proteolysis. Cell *91*, 447–456.

Wendler, P., Ciniawsky, S., Kock, M., and Kube, S. (2012). Structure and function of the AAA+ nucleotide binding pocket. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1823*, 2–14.

Wolny, M., Batchelor, M., Knight, P.J., Paci, E., Dougan, L., and Peckham, M. (2014). Stable Single α-Helices Are Constant Force Springs in Proteins. J. Biol. Chem. *289*, 27825–27835.

Woo, K.M., Chung, W.J., Ha, D.B., Goldberg, A.L., and Chung, C.H. (1989). Protease Ti from Escherichia coli requires ATP hydrolysis for protein breakdown but not for hydrolysis of small peptides. J. Biol. Chem. *264*, 2088–2091.

Yu, Y., Wan, Y., and Huang, C. (2009). The Biological Functions of NF-κB1(p50) and its Potential as an Anti-Cancer Target. CCDT *9*, 566–571.

Zhang, M., and Coffino, P. (2004). Repeat Sequence of Epstein-Barr Virus-encoded Nuclear Antigen 1 Protein Interrupts Proteasome Substrate Processing. J. Biol. Chem. *279*, 8635–8641.