Title
Mechanism of Protein Unfolding and Polypeptide Translocation by the AAA+ Protease ClpXP

Permalink
https://escholarship.org/uc/item/5kr7t56k

Author
Nyquist, Kristofor Belew

Publication Date
2015

Peer reviewed|Thesis/dissertation
MECHANISM OF PROTEIN UNFOLDING AND POLYPEPTIDE TRANSLOCATION BY THE AAA+ PROTEASE CLPXP

By

Kristofor Belew Nyquist

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

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Andreas Martin, Chair
Professor Carlos Bustamante
Professor Jeremy Thorner
Professor Ahmet Yildiz

Summer 2015
ABSTRACT

Mechanism of Protein Unfolding and Polypeptide Translocation by the AAA+ Protease ClpXP.

by

Kristofor Belew Nyquist

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Andreas Martin, Chair

ATP-dependent proteases exist in all cells and are crucial regulators of the proteome. These biological nanomachines contain a homohexameric molecular motor which generates mechanical force to unfold proteins and translocate unstructured polypeptides into a degradation chamber for proteolysis. Using E. coli ClpXP as a model system for investigating the function of ATP-dependent proteases, I used single-molecule optical tweezers to explore how individual subunits within the hexameric motor convert the chemical energy from ATP hydrolysis into the mechanical work required to perform protein unfolding and polypeptide translocation.

These studies revealed key aspects of motor function that highlight how ATP-dependent proteases are tailored to serve their specific cellular tasks. My analyses indicated that ClpXP translocates in bursts resulting from highly coordinated conformational changes in two to four ATPase subunits and that these bursts of translocation couple to the phosphate release step of the ATP hydrolysis cycle. I found that firing of four ATPase subunits is required to successfully unfold GFP, but that polypeptide translocation robustly occurs even when only two ATPase subunits fire. Because this work pushed the envelope of single-molecule data analysis techniques, I developed new algorithms to detect steps in passive-mode optical tweezers data, where the noise statistics vary in time.

The results discussed in this dissertation improve our understanding of how ClpXP mechanically operates. Strong structural similarities to other ATP-dependent proteases, as well as other ring translocases, suggest that the operating principles I have discerned and describe here may generalize to other biological molecular motors.
This dissertation is dedicated to you, the reader. Thanks for looking!
ACKNOWLEDGMENTS

First, I’d like to acknowledge all of the scientists who’ve inspired me and driven me to become a scientist. I would like to especially thank Andreas Martin for his mentorship through the duration of my PhD. His intellect is as unmatched as his intuition for identifying and attacking interesting scientific questions. I will miss the culture of fun curiosity he cultivates in his lab. I would also like to thank Carlos Bustamante for his mentorship during my thesis work. His excitement for dissecting molecular motor mechanism and for appreciating their similarities as equally as their differences has been an infectious addition to my scientific career. My career as a scientist began at Washington State University in the Department of Physics and Astronomy. I am forever indebted to Doerte Blume for welcoming me into her research group and for spending the time to set the foundation for me to grow into a scientist.

I would like to thank Rodrigo Maillard for his guidance during my first few years as a graduate student as we set out to dissect how ClpXP worked. I’d also like to thank my fellow ClpXP teammates, Maya Sen and Pierre Rodriguez. A special thanks goes to my collaborator Steve Pressé for working with me to improve step-fitting algorithms. Even though our interactions took place primarily through Skype, he had a profound impact on my career and our work together has helped to give me direction in my next steps as a scientist. Of course, I would like to thank the entire Martin lab, especially, those of us who worked together through the early years: Mary Matyskiela, Robyn Beckwith, Eric Estrin, Charlene Bashore, and Chris Padovani. Even though at times our work could get stressful, we always managed to have fun, play foosball, and listen to Kanye.

The last few years have been enriched by many friends, both old and new. From weekend trips to Reno to visit my oldest friend Wes Helander, to random local debauchery with my close friends Michael Souza, Vlad Belyy, and Tim Wendorff, I’ve had fun getting the occasional chance to rest from my thesis work.

The most important thanks goes to Sara Wahlin. She’s been my rock since the beginning and, without her love and support, I would not have had the staying power to finish this dissertation.

Finally, I’d like to thank my parents. They fostered a love of education at a very young age and taught me to always satisfy my curiosity. Most importantly, I’d like to thank them for teaching me to never be afraid of learning. They instilled the confidence, even at a young age, that everything is learnable and that nothing is out of reach if you work hard enough.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
</tr>
<tr>
<td>1 Introduction</td>
</tr>
<tr>
<td>2 Mechanochemistry of ClpX subunits</td>
</tr>
<tr>
<td>3 Intersubunit coordination within the ClpX hexamer</td>
</tr>
</tbody>
</table>
3.2.1 ATPγS dependence of pause density and pause duration reveals high intersubunit coordination during translocation  ........ 28
3.2.2 Translocation heterogeneity is intrinsic to mechanism ....... 32
3.2.3 Effect of [ATP] on burst sizes and dwell durations .......... 34
3.2.4 ClpXP requires four highly coordinated power strokes for successful unfolding of GFP ................................. 37
3.2.5 Hydrolysis-deficient ClpX rings translocate robustly but are marginalized in GFP unfolding .............................. 41

3.3 Discussion  ................................................. 45
3.3.1 Intersubunit coordination determines the ClpX translocation mechanism .................................................. 45
3.3.2 An “internal clock” triggers polypeptide translocation .... 49
3.3.3 Biological relevance of ClpX translocation mechanism .... 51

3.4 Materials and Methods ........................................ 52
3.4.1 Measurements of ATPγS-hydrolysis rate ................. 52
3.4.2 Extracting pauses from ATPγS traces ..................... 53
3.4.3 Motivation for fitting ATPγS-induced pauses to the Hill equation .......................................................... 53
3.4.4 Calculating the fraction of ATPγS molecules bound to the hexamer ......................................................... 53
3.4.5 Step-fitting methodology ...................................... 55
3.4.6 Spatial/temporal resolution of the instrument ............ 55
3.4.7 Identifying the structures of ClpXP-mediated GFP-unfolding intermediates ............................................... 57
3.4.8 Prediction of GFP unfolding time based on burst size and dwell time distributions ...................................... 59
3.4.9 GFP-fluorescence degradation assay ........................ 59

4 Concluding Remarks and Future Directions .................... 61

A Fitting steps in passive-mode optical tweezers experiments .... 65
A.1 Introduction ..................................................... 65
A.2 Algorithm ...................................................... 66
A.2.1 Derivation .................................................. 67
A.2.2 Computational scheme for model selection ................ 72
A.3 Results .......................................................... 73
A.4 Discussion ...................................................... 75
A.5 Methods ......................................................... 77
A.5.1 Simulating exponentially correlated Gaussian noise .... 77

REFERENCES ......................................................... 78
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Protein degradation by ATP-dependent proteases.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Architecture of unfoldase rings.</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Topological constraints necessitate conformational switching.</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>Schematic depiction of dual-trap optical tweezers.</td>
<td>9</td>
</tr>
<tr>
<td>1.5</td>
<td>Experimental geometry of ClpXP optical tweezing assay.</td>
<td>11</td>
</tr>
<tr>
<td>1.6</td>
<td>Model substrates used for single-molecule experiments.</td>
<td>12</td>
</tr>
<tr>
<td>1.7</td>
<td>Polypeptide translocation is step-like.</td>
<td>12</td>
</tr>
<tr>
<td>2.1</td>
<td>Translocation is force-insensitive when ATP binding limits translocation rate.</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Force dependence of Michaelis-Menten parameters.</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Competition with ADP and Pi.</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>Model for how a ClpX subunit’s ATP-hydrolysis cycle is coupled to translocation.</td>
<td>20</td>
</tr>
<tr>
<td>3.1</td>
<td>Rate of ClpXP ATPγS hydrolysis measured by thin-layer chromatography.</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>ATPγS induces long-lived pauses.</td>
<td>30</td>
</tr>
<tr>
<td>3.3</td>
<td>Characteristics of ATPγS-induced pauses.</td>
<td>31</td>
</tr>
<tr>
<td>3.4</td>
<td>Translocation is heterogeneous.</td>
<td>32</td>
</tr>
<tr>
<td>3.5</td>
<td>Heterogeneity is intrinsic to translocation mechanism.</td>
<td>33</td>
</tr>
<tr>
<td>3.6</td>
<td>Effect of ATP concentration on burst size and dwell duration.</td>
<td>35</td>
</tr>
<tr>
<td>3.7</td>
<td>GFP unfolding probabilities.</td>
<td>36</td>
</tr>
<tr>
<td>3.8</td>
<td>GFP unfolds via two distinct intermediates.</td>
<td>38</td>
</tr>
<tr>
<td>3.9</td>
<td>Estimating the contour length of GFP-unfolding intermediates using worm-like chain polymer theory.</td>
<td>39</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>3.10</td>
<td>Mapping contour lengths of each intermediate to the primary sequence of GFP.</td>
<td>40</td>
</tr>
<tr>
<td>3.11</td>
<td>Intermediate ATP concentrations reveal failed β11 extraction by ClpXP.</td>
<td>41</td>
</tr>
<tr>
<td>3.12</td>
<td>Schematic depiction of the number and arrangement of the various ClpX mutants considered.</td>
<td>42</td>
</tr>
<tr>
<td>3.13</td>
<td>Rate of translocation decreases as a function of the number of consecutive R subunits.</td>
<td>43</td>
</tr>
<tr>
<td>3.14</td>
<td>Representative traces showing pausing behavior observed during translocation by ClpX sensor-2 mutants.</td>
<td>44</td>
</tr>
<tr>
<td>3.15</td>
<td>Pause-free velocity of translocation remains constant as a function of the number of consecutive R subunits.</td>
<td>45</td>
</tr>
<tr>
<td>3.16</td>
<td>The density of pauses during translocation increases as a function of the number of R subunits.</td>
<td>46</td>
</tr>
<tr>
<td>3.17</td>
<td>Changing the arrangement of R subunits leads to full translocation recovery.</td>
<td>47</td>
</tr>
<tr>
<td>3.18</td>
<td>Even single R-subunit ClpX rings are severely compromised in GFP unfolding.</td>
<td>48</td>
</tr>
<tr>
<td>3.19</td>
<td>Schematic depiction of intersubunit coordination.</td>
<td>50</td>
</tr>
<tr>
<td>3.20</td>
<td>Step-like pattern clearly observed during translocation.</td>
<td>55</td>
</tr>
<tr>
<td>3.21</td>
<td>Step-fitting methodology.</td>
<td>56</td>
</tr>
<tr>
<td>3.22</td>
<td>Lifetime of second intermediate is independent of whether or not the first intermediate was resolved.</td>
<td>58</td>
</tr>
<tr>
<td>A.1</td>
<td>Fitting steps using SIC to a trace with identically and independently distributed Gaussian noise is robust.</td>
<td>67</td>
</tr>
<tr>
<td>A.2</td>
<td>Fitting steps using SIC to a trace with identically, but not independently distributed Gaussian noise fails impressively.</td>
<td>68</td>
</tr>
<tr>
<td>A.3</td>
<td>Depiction of SICP algorithm.</td>
<td>72</td>
</tr>
<tr>
<td>A.4</td>
<td>SICP effectively detects steps in traces with non-stationary noise.</td>
<td>74</td>
</tr>
<tr>
<td>A.5</td>
<td>Histograms of step sizes detected in traces with non-stationary noise.</td>
<td>75</td>
</tr>
<tr>
<td>A.6</td>
<td>SICP effectively detects steps in traces with a small number of data points.</td>
<td>76</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 ATP-dependent proteases maintain the proteome

All cells rely on ATP-powered proteolytic machines to selectively degrade targeted protein substrates for quality control and regulation (Sauer and Baker, 2011). These molecular devices share a common architecture, consisting of a ring-shaped protein unfoldase and a barrel-shaped compartmental peptidase. The unfoldase ring is composed of six AAA+ subunits (ATPases associated with various cellular activities) (Hanson and Whiteheart, 2005; Neuwald et al., 1999) that serve as a motor, converting the energy from ATP hydrolysis into the mechanical work required to unfold and translocate protein substrates through its central pore. The unfoldase docks to one or both axial faces of the barrel-shaped peptidase, which contains proteolytic active sites sequestered in an internal chamber. Access to these active sites is restricted by two narrow axial pores with a diameter too small to allow entry of even the smallest folded proteins. The central pore of the docked unfoldase aligns with the pore of the peptidase, enabling the ATPase ring to deliver substrates into the peptidase chamber for degradation.

1.1.1 General trajectory of energy-dependent protein degradation

The protein degradation pathway follows a similar trajectory for all ATP-dependent proteases (Figure 1.1). Substrate specificity is usually determined through placement of degradation tags, or degrons. Degrons can be intrinsic peptide sequences that are hidden within a correctly folded protein and become exposed only upon protein damage or misfolding, for instance, when the bacterial cell is subjected to heat shock (Truscott et al., 2011). Degrons can also be appended to a protein during certain checkpoints, for example, in the ssrA-tagging system for stalled protein synthesis. In this pathway, the ribosome is rescued by a modified transfer-RNA that includes a message to co-translationally append the 11-residue ssrA degron to the C-terminal end of the nascent chain (Giudice and Gillet, 2013). However, not all degrons are small peptide sequences. For instance, the eukaryotic 26S proteasome recognizes protein substrates via polyubiquitin chains covalently attached to surface-exposed lysines (Finley, 2009). Recognition of degrons can occur in many different ways: the degron can bind directly to the pore of the unfoldase (Martin et al., 2008a), to an auxiliary site somewhere else on the protease (Matyskiela and Martin, 2013), or to a cofactor.
Figure 1.1.: Protein degradation by ATP-dependent proteases. A protein substrate is specifically recognized by the unfoldase and engaged by residues located in the central pore. Downstream ATP-dependent steps lead to unfolding of the substrate and subsequent translocation of the unfolded polypeptide through the central pore of the unfoldase into the degradation chamber of the peptidase for proteolysis.

that in turn delivers the substrate to the protease (Kirstein et al., 2009; Inobe and Matouschek, 2008; Baker and Sauer, 2006).

Once bound to the protease, a protein substrate must become engaged with the translocation machinery in the central pore of the unfoldase to allow mechanical unraveling of folded structures and threading of the polypeptide into the peptidase. Cycles of ATP hydrolysis lead to conformational changes within the ATPase subunits of the ring, generating a vectorial force to propel the substrate through the pore and unravel folded domains that are too large to pass (Maillard et al., 2011; Aubin-Tam et al., 2011).

1.2 Subunit conformational changes drive substrate translocation via pore loop contacts

Loops protrude from every ATPase subunit into the central pore (pore-1 loops). These loops are at least in part responsible for transmitting the ATP-dependent conformational changes of the unfoldase ring to the protein substrate (Wang et al., 2001; Park et al., 2005; Martin et al., 2008b; Koga et al., 2009; Zhang et al., 2009). For all AAA+ protein translocases, the pore-1 loop contains a highly conserved aromatic residue, and crosslinking experiments with the homohexameric ClpX unfoldase from *E. coli* have established that its loop tyrosine directly interacts with polypeptide substrates (Martin et al., 2008b). Mutation of this tyrosine to alanine or phenylalanine has been shown to change the rate of ATP hydrolysis and severely affect the speed as well as efficiency of protein unfolding and translocation (Martin et al., 2008b; Siddiqui
Interestingly, the nucleotide state of the subunit carrying the pore-1 loop mutation significantly affected the observed phenotype. When analyzing ClpX mutant hexamers with combinations of active and inactive subunits, loop mutations exhibited more deleterious effects when placed in ATP hydrolyzing versus hydrolysis-deficient subunits (Martin et al., 2008b), consistent with a model where the subunit’s hydrolysis cycle drives loop motion and thus substrate translocation.

The interaction between pore-1 loops and polypeptide substrates must be surprisingly promiscuous, given the many cellular proteins with different amino-acid sequences that need to be grabbed and translocated. In principle, the loop could sense regularities of the peptide bond or even nonspecifically engage the side chains of individual residues, much like the teeth on the chain ring of a bicycle grip the chain. Remarkably, the apparent substrate indifference of AAA+ proteases during translocation extends well beyond the requirements necessary to carry out their natural protein translocation tasks. Studies have shown that polypeptide translocation can occur not only from either terminus, but also irrespective of D- or L-chirality, and with little regard for side-chain chemistry or peptide-bond spacing (Barkow et al., 2009; Lee et al., 2001). However, the side-chain composition and thus complexity or “slipperiness” of a particular substrate segment has been found to affect the probability of unraveling subsequent folded domains with high thermodynamic stability (Too et al., 2013). Certain proteins, such as the transcription factors NFκB, Spt23, Mga2, Gli2, Gli3, and others, contain internal degradation-stop signals with low-complexity sequences that reduce the grip of the unfoldase and result in partial degradation and the release of protein fragments with new functions (Rape and Jentsch, 2002; Palombella et al., 1994; Schrader et al., 2011). It remains unclear exactly what particular polypeptide feature the pore-1 loops grab on to and hopefully future studies will answer this question.

Regardless of how the pore-1 loop interacts with polypeptide, it seems clear that its movement is coupled to the ATP-hydrolysis cycle of the ring subunits. Given the hexameric architecture of the ring, understanding the coordination among subunit-hydrolysis cycles would thus provide critical insights into the mechanism of translocation. Three general classes of inter-subunit coordination for unfoldases have been previously proposed: sequential, concerted, and stochastic firing (Lyubimov et al., 2011). In the sequential model, the hydrolysis cycle proceeds successively in a spatial and temporal order around the ring. In the concerted model, all subunits complete individual steps of the hydrolysis cycle at the same time. Finally, in the stochastic model, each subunit progresses through its hydrolysis cycle independently of the others. Using single-chain variants of ClpX in which all subunits were covalently linked, it has been shown that diverse arrangements of active and inactive subunits still support protein unfolding and polypeptide translocation, although at reduced rates (Martin et al., 2005). These results immediately rule out strictly sequential or concerted models, because they violate the tight synchrony required for such mechanisms. Although these results generally favored a more stochastic model, the ATP-hydrolysis and substrate-translocation rates did not simply correlate with the number of active
Figure 1.2.: Architecture of unfoldase rings. (a) Organization of a single ClpX ATP-binding pocket showing the position of characteristic functional motifs (PDB ID 4I81). The side-chain positioning of the sensor-2 arginine was not resolved in the original crystal structure and has been modeled here for clarity. The inset shows the location of the binding pocket at the interface between two subunits, highlighting that both subunits contribute active-site residues. (b) Hexameric structure of ClpX shows two major conformational classes. Subunits labeled U (for “unloadable”) are in an open conformation leading to a destroyed ATP-binding site (red X). Subunits labeled L (for “loadable”) are in a closed conformation and form an intact ATP binding pocket (green check). Different subunits have distinct colors. One “rigid body” between a small AAA+ domain and the large AAA+ domain of the clockwise-next neighbor is indicated by a dashed outline. (c) Structural alignment of the large domains of every subunit reveals a nucleotide-dependent hinge-like motion of the same subunit’s small domain about its N-linker. (d) Structural alignment of the large domains of every subunit shows the existence of a rigid binding interface with the small domain of the previous subunit. Subunit colors for (c) and (d) are as shown in (b).

subunits, but showed a dependence on the arrangement of nucleotide states in the ring, indicating a certain degree of inter-subunit coordination (Martin et al., 2005).

1.3 Structural properties of the ClpX ring constrain modes of operation

The architecture of unfoldase rings limits their possible modes of operation. To understand the origins of these constraints, we first focus on the structure of the individual ATPase subunits.
1.3.1 Active site organization

Each subunit contains at least one AAA+ module, which is composed of a large AAA+ domain and an α-helical small AAA+ domain, connected by a covalent linkage known as the N-linker (Smith et al., 2004). The nucleotide-binding pocket is formed by conserved motifs located at the interface between the large and small AAA+ domains (Walker A, Walker B, sensor I, sensor II, and Box VII; see Figure 1.2A) (Erzberger and Berger, 2006). Interestingly, nucleotide hydrolysis requires an arginine residue that is donated by the large AAA+ domain of the neighboring subunit. The presence of this “arginine finger” indicates that the hydrolysis capability of the nucleotide-binding site requires oligomerization. Although I omit a detailed discussion of the AAA+ ATP-binding pocket here, I refer the curious reader to more thorough reviews of its structure and function (Hanson and Whiteheart, 2005; Wendler et al., 2012).

1.3.2 Hinge-like motion of individual subunits is coupled

Because ATP hydrolysis requires oligomerization, the structural organization of the hexameric ring provides insight into the conformational degrees of freedom that are important for function. Crystal structures of the bacterial unfoldase ClpX from *E. coli* have revealed a ring architecture with distinct structural asymmetry, despite being composed of six identical ATPase subunits (Glynn et al., 2009; Stinson et al., 2013). Two major classes of subunits were observed within the hexamer (Figure 1.2B). For one class, termed “loadable” (L), the relative orientation of the large and small domains is compatible with an intact nucleotide-binding pocket. In the other class, termed “unloadable” (U), an 80° rotation about the N-linker between the domains destroys the nucleotide binding pocket (Figure 1.2C). An arrangement of L/U/L/L/U/L was observed for the majority of crystallized ClpX variants. The stoichiometry of four L subunits within the hexamer is consistent with binding studies showing that a maximum of four nucleotides can occupy the ring at saturation (Hersch et al., 2005). Intriguingly, the interface between the small domain of a given subunit and the large domain of its clockwise-next neighbor, as viewed from the top of the ring, is invariant regardless of whether the subunits are “loadable” or “unloadable”, leading to the formation of six “rigid bodies” (Figure 1.2D). The static subunit interfaces persist during protein unfolding and translocation (Glynn et al., 2012), indicating that substrate translocation may be driven by nucleotide-dependent movements of the rigid bodies about their connecting N-linker “hinges”.

Whether or not the structural asymmetry observed for ClpX extends as a general feature of AAA+ unfoldases remains to be established. Crystal structures of other unfoldases such as HslU and Lon show a more symmetric ring structure in which all subunits adopt a similar conformation (Bochtler et al., 2000; Cha et al., 2010). These symmetric structures may imply a different mechanism of function; however, they seem at odds with existing biochemical data reporting that the hexameric rings of
HslU, PAN, and ClpX unfoldases can only bind four nucleotides at saturation (Hersch et al., 2005; Yakamavich et al., 2008; Horwitz et al., 2007). It is possible that these structures of HslU and Lon represent nonfunctional conformations or apo states in the absence of nucleotide and/or substrate. Symmetry breaking for these unfoldases could arise after ATP loads on the ring or substrate enters the pore, as evidence for both of these modes of induced conformational changes is emerging. Crystal structures of FtsH in the apo- and nucleotide-bound states show that the unfoldase ring breaks from a six-fold symmetry to a pseudo two-fold symmetric structure comparable to the ClpX hexamer (Bieniossek et al., 2009). In a similar manner, recent cryo-electron microscopy (cryo-EM) structures of the eukaryotic 26S proteasome in the absence and presence of protein substrate show two distinct conformations (Lander et al., 2012; Beck et al., 2012; Matyskiela et al., 2013). Comparison of these two conformations reveals that substrate engagement triggers the formation of rigid bodies reminiscent of those observed in the ClpX hexamer, suggesting that substrate translocation by the proteasome also relies on constrained motions of rigid bodies about their N-linker hinges.

1.3.3 Rigid bodies provide allosteric coupling for inter-subunit coordination

Due to the rigid bodies formed between neighboring ATPase subunits and their limited degrees of freedom imposed by the connecting linkers, motions of a single subunit likely also cause global conformational flexing of the entire ring (Figure 1.3A). Direct observation of this behavior has been possible using a technique called tmFRET, which relies on the distance-dependent quenching of a fluorescent dye attached to a protein within nanometer-scale proximity to a bound transition-metal ion (Taraska et al., 2009). By carefully choosing the placement of the fluorescent-dye probe within a single ClpX subunit, the tmFRET signal could report on either the binding of nucleotide-transition-metal complex or the conformational state of the labeled subunit within an active hexamer (Stinson et al., 2013). Surprisingly, the nucleotide occupancy of a particular subunit did not strictly correlate with its conformational state, implying that the conformation of empty subunits can be influenced by nucleotides bound somewhere else in the ring (Figure 1.3B). This result suggests a certain degree of coordination among subunits in their ATP-hydrolysis and substrate-translocation cycles. A recent *in silico* study arrived at a similar conclusion purely by enforcing the structural constraints described above (Hwang and Lang, 2013). By simulating faux ClpX hexamers composed of same-type subunits (L or U) with obligatory rigid interfaces, it was discovered that the resulting structures adopt a helical ‘open lockwasher’ conformation. Only when mixtures of L- and U-type subunits were allowed to coexist, it was possible to recover the closed-ring architecture known to be present during ClpX operation (Glynn et al., 2012). Likewise, the AAA+ motor domain of cytoplasmic dynein appears to also require mixtures of subunits in different conformations to form a closed ring (Hwang and Lang, 2013), and increasing
Figure 1.3.: Topological constraints necessitate conformational switching. (A) The rigid interfaces between neighboring subunits as well as the closed-ring topology cause conformational changes of a single subunit to also change the conformation of other subunits. The conformational coupling of subunits throughout the hexamer is indicated by a zigzag circle. (B) ATP binding to a subunit results in a hinge-like motion of the small domain about its N-linker (see Figure 1.2C). Other subunits within the ring are forced to switch their conformation (orange-colored subunit changes green) and, equivalently, change their nucleotide binding affinity (square binding site changes its shape). The spatial order of conformational switching and the detailed conformation of switched subunits remain unclear. This illustration depicts the simplest case involving adjacent subunits, with the remainder of the ring blurred from view.

Experimental evidence for conformational switching dynamics in the dynein motor is emerging from structural (Schmidt et al., 2012) as well as mechanistic studies (Huang et al., 2012). Indications of a conformational switching requirement to form closed, functional rings has also been observed in the asymmetric crystal structure of the related RecA-type T7 gene 4 helicase (Singleton et al., 2000). These observations illustrate that strong structural constraints do not only apply to AAA+ protein unfoldases like ClpX but also extend to other ring-shaped motors with significantly different functions. However, it appears that HslU might be an exception (Hwang and Lang, 2013).

1.4 ClpX generates mechanical force for unfolding and translocation

During protein degradation by ClpXP, the ClpX motor exerts mechanical forces to unfold and translocate selected substrates into the ClpP peptidase for proteolysis (Maillard et al., 2011; Aubin-Tam et al., 2011). These processes are kinetically distinct, but have proven difficult to separate using traditional biochemical approaches. Single-molecule optical tweezers provide a straightforward means to temporally segregate the unfolding activity of ClpXP from its translocation activity. Optical tweez-
ers also enable application of assisting or resisting forces to perturb motor activity, providing a natural means for studying the mechanical functions of ClpX.

1.4.1 Dual-trap optical tweezers: form and function

For the majority of experiments described in this thesis, dual trap optical tweezers were used. In the subsequent section, a detailed discussion regarding the ClpXP assay will be provided. Here, a broad overview of optical tweezers will be presented with specific focus placed on the coordinate system of dual-trap optical tweezers, as well as the optical setup for building one, including excellent sources for readers who aim to build their own instrument.

Shown in Figure 1.4A, dual-trap optical tweezers enable precise measurements of distance at molecular length-scales. The physical basis for the trapping force depends on the size of the trapped object compared to the wavelength of laser light used (Lang and Block, 2003; Harada and Asakura, 1996; Mazolli et al., 2003). For many optical trapping experiments, including those described in this thesis, proteins are bound to large polystyrene beads, which serve as the trapped objects. Thus for dual-trap tweezers we can define the following coordinate system. Two beads, each of radius $r_1$, are trapped in laser traps separated by a known distance $\Delta x_T$. With a polymer tethered between the beads, an entropic force can arise that is strong enough to pull the beads out of their trap centers. The distance between each bead’s center and the center of its respective trap ($\Delta x_{TB,1}$ and $\Delta x_{TB,2}$) is measurable (see below). Thus, the end-to-end extension of the tethered polymer can be directly measured as $\Delta x_T - \Delta x_{TB,2} - \Delta x_{TB,1} - 2r_1$. To a good approximation, the trapping force of the laser trap follows Hooke’s law, i.e. $F_{TB,1} = k_{T1} \Delta x_{TB,1}$, where $k_{T1}$ is the stiffness of the first trap. Each trap’s stiffness can be experimentally determined to high accuracy (see below) and is dependent upon the intensity and wavelength of the trapping laser, as well as the temperature of the experiment and the shape, size, and type of trapped object (Rohrbach, 2005; Malagnino et al., 2002). Information regarding the conceptual basis for how these parameters are measured will be provided as we discuss the optical path of dual-trap optical tweezers along with good references for the more curious reader who seeks more detailed information.

The optical path for building a dual-trap optical tweezers instrument is depicted in Figure 1.4B (red light path). Both traps are formed from a single laser by splitting the light by polarization. This method vastly improves the performance of optical tweezers because all correlated noise cancels (Moffitt et al., 2006; Bustamante et al., 2009b). A collimated 1064nm infrared laser (IRL) is directed through a Faraday isolator (FSI) to prevent laser feedback arising from unwanted reflections off downstream optics. The light is fed through a circuit consisting of a half-wave plate (HWP1) followed by a polarizing beam-splitter (PBS1), enabling the overall trapping-light intensity to be easily attenuated into a beam dump (BD). A second half-wave plate (HWP2) is used to tune the distribution of light polarization, which controls the relative intensity and thus trap stiffnesses of the individual laser traps. The light is
expanded to 1 cm using either a Galilean or Keplerian lens-pair telescope (BEX1). This expansion step is not yet necessary, but does improve safety by spreading out the
Figure 1.4.: (a) Cartoon of dual-trap optical tweezers emphasizing the relevant coordinate system. Spherical beads of known radius ($r_1$) are trapped by the intensity gradient of a tightly focused laser beam. The distance between trap centers is known $\Delta x_T$ as well as the relative distance of each bead from its respective trap center ($\Delta x_{TB,1}$ and $\Delta x_{TB,2}$, respectively). The distance between beads, which we shall call the extension, can be calculated as $\Delta x_T - \Delta x_{TB,1} - \Delta x_{TB,2}$. If a polymer is tethered between the beads, the end-to-end extension of the polymer, $\Delta x_T - \Delta x_{TB,2} - \Delta x_{TB,1} - 2r_1$, can be used to calculate the polymer’s contour length using the worm-like chain model (Bustamante et al., 1994; Marko and Siggia, 1995), given the magnitude of applied external force, the temperature, the polymer’s persistence length, and the end-to-end distance of the polymer (i.e. the extension). (b) Optical path for dual-trap optical tweezers. The two traps depicted in (a) are formed in the sample flow-chamber (FCH). See Section 1.4.1 for detailed description of each component.

Light’s intensity over space. Following the first beam expansion, the light is separated by polarization using a second polarizing beam splitter (PBS2). One polarization is reflected off a fixed-position mirror (FM), while the other polarization is reflected off a tiltable piezo-controlled mirror (PM) before being recombined by a third polarizing-beam splitter (PBS3) back into a single beam. The second beam expansion (BEX2) expands the beam to 2.5 cm in order to slightly overfill the back-aperture of the downstream objective (OBJ1). The telescope also serves to conjugate the plane of the piezo-controlled mirror with the back-focal plane of the objective, enabling tilting of the mirror to result in translation of the focused laser trap within the sample flow chamber (FCH). Thus, the relative distance between laser traps changes as a function of the voltage applied to the piezo-controlled mirror in predictable way that can be calibrated (Bustamante et al., 2009a). A second objective (OBJ2) collects the forward-scattered light, which is split again by polarization (PBS4). The scattered light from each trap is imaged onto distinct quadrant photodiodes (QPD1 and QPD2). The pattern of scattered light on the quadrant photodiodes varies as a function of the distance of the trapped bead to the center of its trap (Gittes and Schmidt, 1998; Bustamante et al., 2009a). A light microscope (LED and CCD, blue light path) configured for Koehler illumination (Koehler, 1894) is used to image the bead sample for bead-trapping as well as piezo-mirror voltage-to-distance calibration (Bustamante et al., 2009a). Trapped beads experience constrained Brownian motion as a result of the external trapping force. The power spectrum of the constrained Brownian motion can be used to calculate the stiffness (or spring constant) of each trap-bead pair, enabling conversion of distance-from-trap-center into applied-external force (Bustamante et al., 2009a). For a good discussion of optical tweezers calibration, see the article by Shaevitz (Shaevitz, 2006).
1.4.2 The ClpXP optical tweezers assay

The structural picture of the unfoldase ring painted above provides a valuable perspective to interpret single-molecule ClpX experiments. Remarkably, the use of dual-trap optical tweezers enables the direct monitoring of ClpX-mediated substrate processing at a sensitivity that resolves separate phases of the translocation cycle (Maillard et al., 2011; Aubin-Tam et al., 2011).

As shown in Figure 1.5, the biotinylated ClpXP complex is bound to a streptavidin-coated bead and the digoxigenin-labeled model substrate (Figure 1.6) is bound to an anti-digoxigenin-coated bead via a 3 kb long dsDNA handle (see Section 2.4 for details regarding protocols for protein purification, modification, and attachment chemistries). Each bead is held in a separate trap. The experiment begins when the beads are both in close proximity and ClpXP engages the C-terminus of the ssrA-tagged substrate (Figure 1.5, cartoon left). The dynamics of unfolding and translocation can thereby be observed as changes in the motor position along the protein substrate. Unfolding of three-dimensional structures of various substrates results in a rapid increase in substrate extension (Figure 1.5, cartoon middle) that subsequently gradually decreases as the unfolded polypeptide chain gets translocated into the protease (Figure 1.5, cartoon right). See Section 2.4 for detailed protocols and methods related to these experiments.
Figure 1.6.: Model substrates used for single-molecule experiments. (a) Double GFP substrate consists of an N-terminal eGFP followed by two titin I27\textsuperscript{CM} domains, another eGFP, a third I27\textsuperscript{CM} domain, and a C-terminal ssrA degradation tag. (b) Single GFP substrate consists of an N-terminal eGFP followed by four I27\textsuperscript{CM} domains, and a C-terminal ssrA degradation tag. See Section 2.4.3 for purification strategy.

Figure 1.7.: Polypeptide translocation is step-like. Translocation occurs via discrete steps composed of a long-lived dwell phase (horizontal red arrow indicates duration) and a rapid burst phase (vertical red arrow indicates spatial size of burst). The high spatial and temporal resolution afforded by optical tweezers reveals that the overall translocation process is fundamentally composed of a step-like pattern, consisting of a dwell phase, during which the ClpX position on the substrate remains fixed, and a burst phase, in which ClpX rapidly translocates a certain length of polypeptide through its central pore (Figure 1.7).
1.4.3 Mechanochemical characterization of ClpXP

In this dissertation, I explore how individual subunits within the ClpX ring convert the chemical energy released from ATP hydrolysis into the mechanical work required to perform protein unfolding and polypeptide translocation. In Chapter 2 I present results that strongly suggest that each subunit translocates substrate concomitantly with the chemical step of phosphate release.

Because each ClpX subunit does not act in isolation, but instead within the context of the activities of neighboring subunits, in Chapter 3 I explore how the subunits coordinate their ATP hydrolysis cycles in order to achieve robust polypeptide translocation and protein unfolding activity. The experiments I describe begin broadly and then narrow down in focus as the results point to more specific models of intersubunit coordination. I begin by presenting experiments using the slowly-hydrolyzable ATP analog, ATP\(_{\gamma}S\) to “throw sand in the gears” and observe how the motor responds to stochastically inactivated subunits. We then look more closely at individual cycles of translocation and discover that a kinetic step distinct from ATP binding to at least some subunits controls the timing of the firing cycle, with plasticity regarding the number of ATP-bound subunits. Finally, we consider some preliminary results using different numbers and orders of inactivated subunits to address the preferred binding and firing order of the ClpX ring.

The methods required to analyze the data from these experiments pushed the envelope of data analysis techniques and algorithms that currently exist in the field of single-molecule biophysics. In Appendix A I present new techniques for analyzing optical tweezers data, specifically tailored for experiments performed in passive-mode, where the statistics of the data are non-stationary in time.

This dissertation addresses key questions in the field and provides a strongly-supported model for how the ClpXP degrades stably-folded protein substrates.
2. Mechanochemistry of ClpX subunits


* indicates that first three authors equally contributed

2.1 Introduction

ATP-dependent proteases of the AAA+ family play crucial roles in the maintenance of cellular protein homeostasis, including the clearance of misfolded, aggregated, or damaged proteins, as well as the disassembly of large macromolecular complexes (King et al., 1996; Hanson and Whiteheart, 2005; White and Lauring, 2007). These multimeric molecular machines encounter unique chemical and mechanical challenges during their task of protein unfolding and polypeptide translocation. First, the motor must translocate along a heterogeneous and flexible polypeptide track composed of amino acids with highly diverse chemical and physical properties. Second, the motor must unravel a diverse array of folded protein domains with a range of stabilities that represent mechanical obstacles along the track. Understanding how ATP-dependent proteases perform these tasks will shed light on their general principles of operation and their mechanochemistry, specifically how they convert chemical energy from ATP hydrolysis to mechanical work and how individual subunits are coordinated.

A first step in understanding the operating principles of a molecular motor is to dissect its mechanochemistry, i.e. how it couples the chemical reaction of ATP hydrolysis to its mechanical operation (Keller and Bustamante, 2000). While it has long been understood that ClpX catalyzes the processes of protein unfolding and polypeptide translocation, recent work had established that ClpX operates as a power-stroke motor (Maillard et al., 2011; Aubin-Tam et al., 2011), directly converting the chemical energy released during ATP hydrolysis into the work required to mechanically drive protein unfolding and polypeptide translocation.

In this chapter, I will specifically address questions related to the coupling of an individual ClpX subunit’s ATP hydrolysis cycle with the overall mechanical cycle of the ClpX ring. We utilized single-molecule optical tweezers, which allowed us to probe the motor’s mechanochemical coupling by applying external forces while simultane-
ously perturbing the chemical transitions of the ATPase cycle. This approach enabled us to examine the force-dependence of translocation as different chemical steps in the ATPase cycle were made rate-limiting, allowing us to draw conclusions regarding which step in the ATP hydrolysis cycle couple to the mechanical, force-generation step of the motor.

2.2 Results

To probe the relationship between the generation of mechanical force and the ATPase cycle of the ClpX motor, which includes initial ATP binding, tight binding, hydrolysis, and the release of ADP and inorganic phosphate (see scheme in Figure 2.1A), we studied ClpXP in the presence of various concentrations of ATP, ADP, and inorganic phosphate (Pi). First, we explored the effect of the ATP concentration on translocation and thereby focused solely on the titin I27\textsuperscript{CM} regions of the fusion substrate (Figure 1.6) in order to minimize potential effects of amino acid sequence differences between titin I27\textsuperscript{CM} and GFP. Translocation was punctuated by rare pauses that are typically longer than 1 to 2 seconds (see Section 2.4.7). Pauses were previously shown to be in kinetic competition with translocation and therefore off the main translocation pathway (Maillard et al., 2011). Pause-free translocation followed a general Michaelis-Menten behavior as a function of ATP concentration, $v = \frac{V_{max}[ATP]}{K_m + [ATP]}$ (Figure 2.1B and C).

To determine whether ATP binding is the force-generating step, we examined how the translocation velocity depends on force at various ATP concentrations. If ATP binding is coupled to force generation, then conditions in which ATP binding becomes rate limiting should make the conformational changes that drive translocation also rate limiting. In this case, the motor velocity should be highly sensitive to the applied external load. In contrast, if motor velocity is insensitive to the applied external force when ATP-binding is rate limiting, then ATP-binding cannot be the ATP hydrolysis step that couples to the power-stroke. We found that the translocation velocity of ClpX is largely insensitive to opposing mechanical forces at low ATP concentrations (Figure 2.1D, purple symbols). In contrast, at saturating ATP concentrations ($\geq 500 \ \mu\text{M}$) and opposing forces between 12 and 20 pN, the force-generating step has become rate limiting and ClpX translocation is force-sensitive (Figure 2.1D, blue symbols). These results clearly indicate that ATP binding does not power substrate translocation.

We then analyzed the force dependence of the Michaelis-Menten parameters $V_{max}$ and $K_m$ to determine where else the force-generating step may be located in the nucleotide-hydrolysis cycle. The Michaelis-Menten fits to our data revealed that both $V_{max}$ and $K_m$ decrease with the applied load (Figure 2.3A), but the $K_m/V_{max}$ ratio remains relatively force insensitive (Figure 2.3B). A force-independent $K_m/V_{max}$ is consistent with our conclusion that ATP binding does not power translocation and indicates that the force-generating step must occur after the first irreversible transition connected to ATP binding (Chemla et al., 2005; Keller and Bustamante, 2000;
Figure 2.1.: Translocation is force-insensitive when ATP binding limits translocation rate. (a) General scheme depicting a motor (M) that binds to one ATP molecule (T), undergoes a tight binding transition, and hydrolyzes ATP, followed by the release of inorganic phosphate (Pi) and ADP. (b) Representative trajectories for translocation of the I27\textsuperscript{CM} moiety of the fusion substrates measured between 6 and 12 pN at different ATP concentrations with ATP regeneration system. The trajectories are offset for clarity and the dashed, parallel lines show the slope at saturating [ATP] for visual clarity. (c) Pause-free velocity shown as a function of [ATP] at 7.5 pN resisting force. Each data point represents the average of the pause-free velocities from a group of individual trajectories (n=10-50) with SEM error bars. Data were fit to a simple Michaelis-Menten equation with $K_m = 31 \pm 6\mu M$ (SEM) and $V_{max} = 8.7 \pm 0.4$ nm/s (SEM). (d) Pause-free velocity of translocation (mean ± SEM) as a function of external force at 5 mM ATP (blue symbols) and 35 µM ATP (purple symbols).

Visscher et al., 1999). This irreversible transition is most likely the tight binding of ATP, as previously observed in other ring ATPases such as the F1-ATPase and the φ29 DNA packaging motor (Adachi et al., 2007; Chemla et al., 2005; Oster and Wang, 2000).

The results above leave only ATP hydrolysis, ADP release, or Pi release as possible transitions that may couple to the force-generating step (i.e. the conformational change within the ring that drives translocation). ATP hydrolysis is unlikely to power translocation, because the small rotation of the terminal phosphate upon hydrolysis does not provide enough free energy change to explain the maximal work we observe, given the stall force of ClpXP (Oster and Wang, 2000).

Eliminating ATP binding and hydrolysis leaves only product release. To investigate whether the release of ADP or Pi drives the translocation step of ClpX, we conducted classical competition assays, varying the concentrations of ADP and
Pi. The effect on pause-free velocity of increasing the ADP concentration at fixed ATP concentrations obeyed a Michaelis-Menten model of competitive inhibition (Figure 2.3A). The apparent $K_m$ for ATP increased linearly with [ADP] according to $K_m = K_m^0 (1 + [ADP]/K_i)$, whereas $V_{max}$ remained constant (Segel et al., 1975). In contrast, increasing [Pi] from 5 µM to 10 mM did not affect the translocation velocity (Figure 2.3B), indicating that Pi release is a largely irreversible transition.

To discriminate between the possible roles of ADP and Pi release in the mechanism of force generation, we estimated the free energy changes of these events based on their respective dissociation constants and compared them to the maximum work performed by ClpXP. Using an estimated fundamental step size of 1 nm (Glynn et al., 2009; Maillard et al., 2011; Aubin-Tam et al., 2011) and a stall force of at least 20 pN, ClpX subunits perform a near-maximum work of 20 pN * 1 nm of $\Delta G = 4.8 k_B T$ when taking a 1 nm step near stall. Using $K_i = 33$ µM as a dissociation constant for ADP, the change in free energy from ADP release is $\Delta G_D \sim 1.8 k_B T$ (at [ADP] = 5 µM) and thus is insufficient to account for the work performed by ClpXP. In contrast, phosphate release is essentially irreversible even at concentrations as high as 10 mM. Therefore, the dissociation constant for phosphate release must be $K_d \gg 10$ mM, with a corresponding change in free energy of $\Delta G_P \gg 7.6 k_B T$ (for [Pi] = 5 µM). Phosphate release would thus provide sufficient energy to power the work produced by ClpXP in every translocation step and is the most likely candidate for

Figure 2.2.: Force dependence of Michaelis-Menten parameters. (a) $K_m$ (blue) and $V_{max}$ (green) plotted against force. (b) $K_m/V_{max}$ ratio plotted as a function of force. Error bars indicate propagated error from each $K_m$ and $V_{max}$ fit (SEM).
2.3 Discussion

These results suggest a mechanochemical model for ClpXP that identifies the force-generating step in the chemical cycle of ATP hydrolysis (Figure 2.4). This work reveals that all transitions reversibly connected to ATP binding up to and including the first irreversible step are not involved in force generation. Furthermore, the first irreversible transition in other molecular motors has been identified as the tight binding of ATP (Adachi et al., 2007; Chemla et al., 2005). Because tight ATP binding is apparently not rate limiting during translocation, its rate constant ($k_{TB}$) can be estimated by using a lower bound that corresponds to the translocation rate of the motor, $k_{cat} = 9 \text{s}^{-1}$. The reverse transition from tight to loose binding, $k_{-TB}$, is given as the inverse of the mean ATPγS-pause duration, $0.6 \text{s}^{-1}$. Thus, we obtain a $k_{TB}/k_{-TB} = 15$ and a corresponding free energy change associated with tight binding $\Delta G_{TB} > 2.7 k_B T$. The tight binding of ATP can therefore be considered the first irreversible transition following ATP binding. Previous work excludes the possibility of ATP hydrolysis being coupled to the force-generating step (Oster and Wang, 2000; Chemla et al., 2005; Keller and Bustamante, 2000). Our results exclude ADP release.

Figure 2.3.: Competition with ADP and Pi. (a) $K_m$ (blue) and $V_{max}$ (green) plotted against ADP concentration at 7.5 pN. Error bars indicate propagated error from each $K_m$ and $V_{max}$ fit (SEM). (b) Pause-free velocity of translocation (mean ± SEM) plotted as a function of phosphate concentration [Pi] at 7.5 pN with a fixed [ATP].
Figure 2.4.: Model for how a ClpX subunit’s ATP-hydrolysis cycle is coupled to translocation. An empty ClpX subunit (E, red sphere) binds ATP (T, orange) and undergoes a tight binding of ATP (T*, green). The subunit hydrolyzes ATP to ADP and Pi (DP, blue). Force-generation, or translocation, occurs upon the release of phosphate (D, purple) and ADP dissociates, leaving an empty subunit (E, red sphere) that’s ready to accept another ATP molecule. The ordering of ADP and phosphate release remains to be determined.

from coupling to force-generation and instead favor a model where force-generation occurs upon Pi release. Interestingly, in this aspect, ClpX resembles other members of the ASCE family, such as the φ29 DNA-packaging motor and the F1-ATPase, harnessing Pi release as a force-generating step despite the distinct architectures and functions within this large family of motors (Chemla et al., 2005).

2.4 Materials and Methods

2.4.1 Single-chain ClpX purification protocol

His-tagged single-chain ClpX expression plasmids are available upon request from the Martin lab. The constructs are housed in pACYC-Duet-1 plasmids and expressed in BLR E. coli cells, which are recA knockout strains and thus prevent homologous recombination of individual ClpX subunits. Single-chain ClpX constructs contain a C-terminal BirA-recognition sequence for biotinylation. Cells are grown in Terrific Broth to OD 1.5 at 37°C, then cooled down to 18°C and induced with 1 mM IPTG overnight.

Harvested cells are resuspended in about 20 mL/L culture Ni-washing buffer (NiA, see composition below) with the addition of 1 mM PMSF for protease inhibition, 2 mg/mL lysozyme to aid lysis, and 250 units benzonase (Sigma) to degrade nucleic acid. Cells are lysed by 2 freeze/thaw cycles and sonication for 2 minutes on ice (15 second bursts followed by 1 minute “rests” to limit sample heating).

After lysis, the lysate is centrifuged at 15k rpm for 30 minutes to pellet cell material. The supernatant is bound in batch to NiA-equilibrated Ni-agarose beads (∼1 mL beads per 1 L culture) for ∼45 minutes and washed in batch ∼4 times with
NiA or until the supernatant is no longer blue by Bradford (i.e. until no protein contaminates exist in solution). The beads are loaded into a gravity column and eluted with 250 mM imidazole (NiB buffer).

The ClpX eluate is biotinylated using recombinantly expressed BirA. I usually concentrate ClpX to about 40 µM and add about 1 µM BirA to the eluate in addition to 0.05 mg/mL D-biotin, 100 mM MgCl₂, and 10 mM ATP (keeping the pH about 7.6). The biotinylation mixture is kept at 30° C for about 1 hour before being loaded on to an S200 size-exclusion column (GE HiLoad 16/60 Superdex 200) equilibrated in gel-filtration buffer (GF). ClpX elutes with a peak position at 58 mL. The appropriate fractions are collected and spin-concentrated to desired concentration (for my experiments usually to about 10 µM). Concentration is determined by UV absorbance at 280 nm with an extinction coefficient of 81000 M⁻¹cm⁻¹. Purified ClpX is flash frozen in liquid nitrogen and stored at -80° C.

Buffers:

NiA
20 mM HEPES, pH 7.6
100 mM KCl
400 mM NaCl
20 mM imidazole
10% glycerol
10 mM BME (add right before use)

NiB
20 mM HEPES, pH 7.6
100 mM KCl
400 mM NaCl
250 mM imidazole
10% glycerol
10 mM BME (add right before use)

GF
50 mM Tris, pH 7.5
300 mM KCl
0.1 mM EDTA
10% glycerol
1 mM DTT (add right before use)

2.4.2 ClpP purification protocol

His-tagged ClpP expression plasmids are available upon request from the Martin lab. The construct is housed in pET-Duet-1 plasmids and expressed in DH5α E. coli cells. Cells are grown in dYT broth to OD 0.6 at 37° C and induced with 1 mM IPTG for 2-3 hours at 30° C.

Harvested cells are resuspended in about 20 mL/L culture S-buffer with the addition of 2 mg/mL lysozyme to aid lysis, and 250 units benzonase (Sigma) to degrade
nucleic acid. Cells are lysed by 2 freeze/thaw cycles and sonication for 2 minutes on ice (15 second bursts followed by 1 minute “rests” to limit sample heating).

After lysis, the lysate is centrifuged at 15k rpm for 30 minutes to pellet cell material. The supernatant is bound in batch to S-buffer-equilibrated Ni-agarose beads (~1 mL beads per 1 L culture) for ~45 minutes and washed in batch ~4 times with S-buffer or until the supernatant is no longer blue by Bradford (i.e. until no protein contaminates exist in solution). The beads are loaded into a gravity column, washed with 25 mL of W20 buffer, and eluted with 500 mM imidazole (W500 buffer). If ClpP crashes out (sign of good expression), it can be put back into solution by the addition of small amounts of EDTA (up to 5 mM, but use as little as possible to not chelate the Ni beads).

The ClpP eluate is buffer exchanged into low-salt buffer (Q50) using a PD-10 desalting column and loaded onto a MonoQ ion-exchange column. MonoQ-bound ClpP is eluted with a 20 mL linear gradient from Q50 buffer to Q1000 buffer. The ClpP eluate is buffer exchanged (PD10) into ClpP storage buffer (Clp) and spin concentrated to the desired concentration before being flash frozen in liquid nitrogen and stored at -80°C.

**Buffers:**

- **S-buffer**
  - 50 mM Na-Phosphate, pH 8.0
  - 1 M NaCl
  - 5 mM imidazole
  - 10% glycerol

- **W20**
  - 50 mM Na-Phosphate, pH 8.0
  - 1 M NaCl
  - 20 mM imidazole
  - 10% glycerol

- **W500**
  - 50 mM Na-Phosphate, pH 8.0
  - 1 M NaCl
  - 500 mM imidazole
  - 10% glycerol

- **Q50**
  - 50 mM Tris-HCL, pH 8.0
  - 50 mM KCl
  - 10 mM MgCl$_2$
  - 1 mM DTT (add right before use)
  - 10% glycerol

- **Q1000**
  - 50 mM Tris-HCL, pH 8.0
  - 1000 mM KCl
  - 10 mM MgCl$_2$
1 mM DTT (add right before use)
10% glycerol
Clp
50 mM HEPES, pH 7.5
200 mM KCl
25 mM MgCl$_2$
0.1 mM EDTA
1 mM DTT (add right before use)
10% glycerol

2.4.3 ybbr-eGFP-(I27$^{CM}$)$_4$-His6-ssrA purification protocol

His-tagged model GFP substrates are available upon request from the Martin lab. The constructs are housed in pACYC-Duet-1 plasmids and expressed in BLR E. coli cells, which are recA knockout strains and thus prevent homologous recombination of individual I27 domains. Cells are grown in dYT to OD 0.6 at 30$^\circ$ C, then induced with 0.23 g/L IPTG for 3-4 hours at 30$^\circ$.

Harvested cells are resuspended in about 20 mL/L culture Ni-washing buffer (NiA) with the addition of 1 mM PMSF for protease inhibition, 2 mg/mL lysozyme to aid lysis, and 250 units benzonase (Sigma) to degrade nucleic acid. Cells are lysed by 2 freeze/thaw cycles and sonication for 2 minutes on ice (15 second bursts followed by 1 minute “rests” to limit sample heating).

After lysis, the lysate is centrifuged at 15k rpm for 30 minutes to pellet cell material. The supernatant is bound in batch to NiA-equilibrated Ni-agarose beads (~1 mL beads per 1 L culture) for ~45 minutes and washed in batch ~4 times with NiA or until the supernatant is no longer blue by Bradford (i.e. until no protein contaminates exist in solution). The beads are loaded into a gravity column and eluted with 250 mM imidazole (NiB buffer).

The substrate eluate was carboxymethylated according to previously established methods (Kenniston et al., 2003). The I27 domains were selectively unfolded with 5 M GuHCL at pH 8 (eGFP remains folded while I27(V15P) becomes unfolded) and 100-fold molar excess of iodoacetic acid was added to irreversibly carboxymethylate the normally buried cysteines. This mixture was incubated for 2 hours at 30$^\circ$ C. Alkylation was assessed by monitoring the red-shift in fluorescence of a normally buried tryptophan, which reads out solvent exposure.

As a final purification step, the carboxymethylated substrate mixture was loaded on an S200 size-exclusion column (GE HiLoad 16/60 Superdex 200) equilibrated in gel-filtration buffer (GF). The appropriate fractions were collected and spin-concentrated to desired concentration (for my experiments usually to about 200 $\mu$M). Purified substrate was flash frozen in liquid nitrogen and stored at -80$^\circ$ C.

Buffers:
NiA
20 mM HEPES, pH 7.6
100 mM KCl
400 mM NaCl
20 mM imidazole
10% glycerol
10 mM BME (add right before use)

NiB
20 mM HEPES, pH 7.6
100 mM KCl
400 mM NaCl
250 mM imidazole
10% glycerol
10 mM BME (add right before use)

GF
50 mM Tris, pH 7.5
300 mM KCl
0.1 mM EDTA
10% glycerol
1 mM DTT (add right before use)

2.4.4 Preparation of DNA handles and substrate-handle attachment

The 3 kbp handles that were used to covalently attach substrates were synthesized via PCR from pPROEX-HTa plasmid. The reverse primer for amplification had a 5' Digoxigenin modification for attachment to antidigoxigenin-coated beads.

Primers
BbsI fw
GACGATACCGAAGACAGGTCGTGTTATATCC
Dig rev
(5’ Dig) GAGCGGTATCAGCTCACTCAAAG
The PCR fragment was digested with BbsI to completion and stored at -80° C.

The 3 kb dsDNA handle was attached to the N terminus of the substrate using the ybbr/Sfp system. Sfp was used to covalently link a CoA moeity (modified with a short dsOligo with a BbsI site) to the N-terminus of the substrate. The dsOligo-CoA-substrate could then be ligated to the 3 kb handle via the BbsI complementary sequence.

2.4.5 Data collection

Two different dual-trap optical tweezers with 1064 nm lasers were employed (Moffitt et al., 2006) (see Figure 1.4B for schematic diagram). The unfolded polypeptide contour length was calculated using the worm-like chain model for polymer elasticity. See Section 1.4.1 for detailed information regarding the dual-trap optical tweezers.
coordinate system as well has information and references regarding how raw signals from optical tweezers are converted into relevant physical parameters.

2.4.6 Measurements of ATP-hydrolysis rate

The ATP hydrolysis rate of ClpXP was measured in bulk using an NADH-coupled ATP-regeneration system as previously described (Martin et al., 2008b; Maillard et al., 2011). Assembled hexamers of ClpX (0.3 µM) were mixed with ClpP (1.5 µM) in a ClpX-100 buffer (25 µM HEPES pH 7.6, 20 mM MgCl₂, 100 mM KCl, and 0.5 mM EDTA) containing an NADH-coupled regeneration system (3 U/mL pyruvate kinase, 3 U/mL lactate dehydrogenase, 1 mM NADH, and 7.5 mM phosphoenolpyruvate). The ATP-hydrolysis rate of ClpX was measured both in the presence and absence of 10 µM titinCM-ssrA by monitoring the absorbance of NADH (340 nm) at 30 °. The observed ATPase rates were consistent with previous studies (Martin et al., 2005; Aubin-Tam et al., 2011).

2.4.7 Pause detection and translocation rate calculation

Pauses were removed from traces using a previously described modified step-fitting algorithm (Kalafut and Visscher, 2008; Chistol et al., 2012) with a pause threshold and penalty. The cutoff threshold was calculated by taking three standard-deviations of a gamma distribution fitted to the dwell-time distribution. Translocation rate was calculated as the end-to-end distance of each translocation region \( \Delta x/\Delta t \). To calculate pause-free translocation velocity, the duration of the detected pauses was subtracted from the \( \Delta t \) component.
3. Intersubunit coordination within the ClpX hexamer


* indicates that first three authors equally contributed

3.1 Introduction

Having discovered that the power-stroke that drives translocation couples to the phosphate-release step during the chemical cycle of ATP hydrolysis, we now focus our attention on figuring out how individual subunits coordinate their activities within the ring. Because these machines must encounter unique challenges during their primary task of protein unfolding and polypeptide translocation, we will discover that these machines have adapted a novel mechanism of intersubunit coordination well-suited to their particular cellular job.

Previous mutagenesis studies have suggested that the ClpX subunits contribute additively to substrate processing and that the power stroke for translocation may be generated by ATP hydrolysis in one subunit at a time (Martin et al., 2005), supporting a probabilistic mechanism of ring subunit coordination (Glynn et al., 2009; Stinson et al., 2013). Although our data do support certain aspects of stochasticity, such as in the initial binding of ATP, we find that the power-stroke phase of translocation occurs with a high degree of intersubunit coordination.

In this chapter, I address the following questions: do all ATPase subunits participate during substrate translocation and what is the mechanism of coordination within the hexameric ring? Is the coordination among subunits different during protein unfolding versus processive translocation of an unstructured polypeptide? Which transitions determine the timing of the mechanochemical cycle? To address these mechanistic questions, we again utilized single-molecule optical tweezers. These studies provided us with several important findings. We determined that to stall the motor during polypeptide translocation ATP hydrolysis in at least three of the six subunits must be inhibited. We find that a process not coupled to ATP binding sets the dwell duration between translocation bursts and that the burst size depends on
the number of hydrolyzing ClpX subunits in the hexamer. This number distributes between two, three, and four subunits and their relative occurrence changes as [ATP] is varied from Km to saturation. During a burst, the near-simultaneous firing and translocation by two, three, or four subunits occurs in a coordinated fashion before the hexamer starts a new mechanochemical cycle. We find that these highly coordinated power strokes occur upon phosphate release and that they play a crucial role in the ability of ClpXP to denature kinetically stabilized protein substrates like GFP. Contrary to previously proposed probabilistic models (Martin et al., 2005; Glynn et al., 2009; Stinson et al., 2013), our results establish a high degree of coordination between ATP-bound subunits in the ClpX hexamer. Although there does seem to be stochasticity in the order in which subunits fill with ATP and the number of subunits that bind ATP, we observed a high degree of coordination during the power-stroke that drives translocation. Interestingly, using catalytic mutants of ClpX in which specific numbers and arrangements of subunits are unable to hydrolyze ATP, we discovered a surprising distinction between translocation activity and protein-unfolding activity, demonstrating that the requirements for translocation significantly deviate from the requirements for protein unfolding. ClpX seems to employ a novel mechanism of translocation that significantly deviates from canonical motor mechanisms, demonstrating how specialized molecular machines have been optimized to carry out their specific tasks.

3.2 Results

Effective unfolding and/or translocation of protein substrates by ClpX may require the participation of multiple subunits in the hexamer (Martin et al., 2008b). We therefore sought to characterize how individual subunits coordinate their ATP hydrolysis activities around the ring.

3.2.1 ATPγS dependence of pause density and pause duration reveals high intersubunit coordination during translocation

We slowed down the hydrolysis in a given subunit by using the ATP analog ATPγS, which ClpX hydrolyzes ~90 times slower than ATP (Figure 3.1) and determined how binding of this analog affected the polypeptide translocation by the remaining ATP-bound subunits in the ring. We held [ATP] fixed at 500 µM and varied [ATPγS] from 0 to 250 µM. In the presence of ATPγS, we observed pauses longer than 1 s, which were extremely rare in the presence of ATP alone, and we attributed these long pauses to ATPγS-bound ClpX subunits (Figure 3.2A). Furthermore, the relationship between the average translocation velocity and the ATPγS concentration decreased in a nonlinear fashion due to these long pauses. The trend fit well to a Hill equation with a Hill coefficient \( n_{\text{ATPγS}} = 1.5 \pm 0.3 \) (SEM) (Figure 3.2B), indicating
Figure 3.1.: Rate of ClpXP ATPγS hydrolysis measured by thin-layer chromatography. (a) Thin-layer chromatography assay of 35S-labeled ATPγS hydrolysis by ClpXP. 35S-ATPγS was incubated with 0.3 µM ClpX and 1.5 µM ClpP (top) or buffer only (bottom) for the time indicated, before being quenched with 2.5 volumes of stop buffer and spotted on TLC plates (see Section 3.4.1). The positions of ATPγS and PO₃S are indicated. (b) Rate of ClpXP-mediated hydrolysis of ATPγS as a function of [ATPγS] in the presence (red) and absence (blue) of I27CM-ssrA. Conditions were as in (a) with the addition of 10 µM I27CM-ssrA, and $k_{cat} = 2.5 ± 0.2$ min⁻¹ ClpX⁻¹ and $K_m = 6.2 ± 2.0$ µM in the absence of I27CM-ssrA. All fitted values are mean ± SEM. (c) Rate of ATP hydrolysis by ClpXP is shown as a function of [ATP] in the presence (red) and absence (blue) of I27CM-ssrA. Fitting the data to the Michaelis-Menten equation provided the values of $k_{cat} = 497 ± 20$ min⁻¹ hexamer⁻¹ and $K_m = 57.2 ± 6.1$ µM in the presence of I27CM-ssrA, and $k_{cat} = 111 ± 13$ min⁻¹ hexamer⁻¹ and $K_m = 58.9 ± 2.0$ µM in the absence of I27CM-ssrA. All fitted values are mean ± SEM.

that more than one ATPγS molecule binding to the ring is necessary to induce a long pause during translocation.
Figure 3.2: ATPγS induces long-lived pauses. (a) Representative trajectories for forces between 6 and 12 pN were measured with increasing [ATPγS] at fixed [ATP]. Trajectories are offset for visual clarity. (b) Translocation rate (mean ± SEM) plotted against [ATPγS], with the fit shown in red.

We observed that the pause density (PD), i.e., the number of pauses per nm of translocated polypeptide, increased with the ATPγS concentration, indicating that the entry into a pause was caused by the binding of ATPγS to the ring (Figure 3.3A, inset). The maximum PD, PD\text{max}, reflects the existence of an ATPγS concentration at which the motor has nearly 100% probability of binding the minimum number of ATPγS molecules required to stall translocation. Accurate pause detection became difficult at high concentrations of ATPγS (greater than \oversim{200-250} \mu M). To calculate PD\text{max}, we plotted the inverse of the ATPγS concentration against the inverse of PD and estimated a PD\text{max} of \oversim{0.5} nm\textsuperscript{-1} (Figure 3.3A). Thus, the motor has a 50% probability of entering a pause at [ATPγS] = 200 \mu M (Figure 3.3A, inset). Using a K\text{m,ATPγS} = 29 \mu M as an upper bound for the ATPγS dissociation constant (because k\text{cat} has little contribution), we calculated that the probability of having three or more ATPγS molecules bound to the ring at [ATPγS] = 200 \mu M is \oversim{0.44} (see Section 3.4.4), very close to the observed 50% probability of the motor entering into an analog-induced pause. We conclude that at least three molecules of ATPγS are required to bind the ring in order to stall translocation and induce a pause. Consequently, intersubunit coordination around the ring does not require the involvement of all ATP-binding-competent subunits.

One of two different processes is likely to dictate the kinetics of exit from a pause: (1) the dissociation or (2) the hydrolysis of ATPγS. To distinguish between these
alternatives, we compared the ATPγS hydrolysis rate with the average duration of the ATPγS-induced pauses. Hydrolysis of a single ATPγS molecule takes approximately 10 s in the presence of I27CM-ssrA (k\text{cat} of \sim 6 \text{ min}^{-1} \text{ hexamer}^{-1}), which is \sim 90 times slower than the rate of ATP hydrolysis under identical conditions (Figure 3.1B and C). Because the observed mean pause durations are significantly shorter than the time for ATPγS hydrolysis (Figure 3.3B), we conclude that they primarily reflect the off-rate of ATPγS. Interestingly, the mean pause durations increased from \sim 1.5 to \sim 2.5 s as the ATPγS concentration was increased from 50 to 250 \mu M (Figure 3.3B). Thus, the exit from a pause takes longer as the motor loads with an increasing number of ATPγS molecules, most likely reflecting the extra time required to eject multiple ATPγS molecules until just two ATPγS remain bound to the ring. Experiments suggest that the off rate of nucleotide, at least in part, depends on the nucleotide-hydrolysis state of neighboring subunits (Stinson et al., 2013). Thus, the non-linearity of the ATPγS-induced pause duration with [ATPγS] could be explained by the overall increased ATPγS occupancy of the ring.
3.2.2 Translocation heterogeneity is intrinsic to mechanism

The observation that ClpX maintains operation even when one or two ATPγS molecules are bound to the ring suggests that the translocation mechanism exhibits plasticity with regard to the filled state of all ATPase subunits. Consistently, as shown in Figure 3.4, ClpXP translocation traces show widely varied features. For example, in Figure 3.4A, homogeneous translocation is observed whereas in Figure 3.4B the motor gradually slows down to a complete halt before translocation near-instantaneously resumes at speeds near maximal values. In Figures 3.4C and D, similar phenotypes are observed that illustrate translocation characteristics between these two extremes.

In general, two simple models could explain the translocation heterogeneity observed in these traces. Since these measurements record the trajectories of single-molecules, our experiments could select for different classes of motors; for instance, some which translocate fast and others that translocate more slowly. Alternatively, the observed heterogeneity could reflect the intrinsic mechanism of translocation. Instead of multiple classes of motors, there could be a single class that's capable of switching between different translocation states, perhaps reflecting different filled-ring states.
Figure 3.5.: Heterogeneity is intrinsic to translocation mechanism. (a) Example trace illustrating the binning process for generating residence-time distributions. The trace was sliced by extension (dashed horizontal lines) and the number of data points was counted and converted into time using the sampling frequency of the data (right). (b) The residence time distribution for all traces obtained under saturating [ATP] (5 mM). Colors within each bar indentify the trace from which that residence time measurement originated.

We distinguished between these models by considering the residence-time distribution of ClpXP translocation. As depicted in Figure 3.5A, each trace was sliced into 10 nm bins along the extension of the polypeptide during translocation. The number of datapoints within each bin provides a proxy from the amount of time that ClpXP resides at that position along the polypeptide track.
The overall distribution of residence times is shown in Figure 3.5B, with each color indicating which trace a particular bin originated from. Clusters of color would indicate different populations of motors. We instead observed stripes of color across the full range of possible residence times, indicating that individual motors indeed display the full range of possible translocation characteristics. Thus the translocation heterogeneity observed in our traces seems intrinsic to the mechanism of translocation. Taken with the results from the ATP-γS titrations, the heterogeneity could arise from translocation occurring during different loaded states of the ring. In order to test this hypothesis, we took advantage of our ability to resolve individual translocation cycles given the relatively high spatial and temporal resolution offered by dual-trap optical tweezers.

3.2.3 Effect of [ATP] on burst sizes and dwell durations

The observation that the motor can maintain its operation even when one or two analog molecules are bound to the ClpX ring suggests that the translocation mechanism exhibits plasticity with regard to the loaded state of the ring. This result led us to investigate in greater detail the mechanism by which subunits communicate with each other around the ring. Previous single-molecule studies showed that ClpXP translocates polypeptide in bursts of 1, 2, and 3 nm (Maillard et al., 2011). Based on the ClpX crystal structure that shows an ~1 nm displacement between the pore-1 loops in the ATP-bound and empty states (Glynn et al., 2009), a 1 nm step was thought to reflect the basic power stroke of a single ClpX subunit (Maillard et al., 2011; Aubin-Tam et al., 2011). Accordingly, the 2 and 3 nm bursts were interpreted as the near-simultaneous firing of two and three ClpX subunits, respectively (Maillard et al., 2011). An important remaining question is what mechanism individual ATPase subunits use to coordinate their translocation activities at any given ATP concentration. To address this point, we analyzed how different ATP concentrations affected the distributions of burst sizes and dwell times (Figure 3.6).

At [ATP] \( \gg K_m \), burst sizes of 2-4 nm were observed at opposing forces between 6 and 15 pN (Figure 3.6B, blue). The distribution of burst sizes has a maximum at ~3 nm, with a correspondingly lower number of ~2 and ~4 nm bursts. In contrast, when ATP binding is near rate limiting ([ATP] = 35 µM), ClpXP translocates mostly in 2 or 3 nm bursts, with 2 nm bursts being most frequent and bursts of 4 nm completely absent (Figure 3.6B, red).

Next, we analyzed the duration of the cycle time (the sum of the dwell and burst phase duration) at various ATP concentrations to better understand the translocation mechanism of this motor. Surprisingly, we observed that the mean cycle time has no apparent dependence on [ATP] (Figure 3.6C). We found that the mean duration of the dwell phase \( < \tau > \) is 350 ± 20 ms in the range between 35 µM and 5 mM (Figure 3.6D). In addition, the duration of the burst phase constitutes to less than ~3 % of the cycle time and has a mean duration of less than ~10 ms. In this range of nucleotide concentrations, the motor translocation rate approximately doubles from
Figure 3.6.: Effect of ATP concentration on burst size and dwell duration. (a) Representative trajectories of ClpXP translocating substrate in 3 nm bursts at 10-14 pN and different ATP concentrations. Raw data were filtered and decimated to 50 Hz (red, blue, and green). Sliding-window t-test fits to the data are shown in black (see Section 3.6). (b) Burst size distribution for ATP concentrations near $K_m$ (red) and saturating (blue) levels. (c) Dwell time distribution for ATP concentrations near $K_m$ (red) and saturating (blue) levels. (d) Mean dwell duration ($\pm$ SEM) plotted against [ATP].

5 to 9 nm/s. Consequently, the observed change in translocation rate is not due to changes in the mean cycle time, but rather to a systematic increase in motor burst size with increasing [ATP].

To obtain insight into the molecular processes that occur during the dwell time, we calculated the kinetic parameter $n_{\text{min}}$ (Schnitzer and Block, 1995; Moffitt et al.,
Figure 3.7.: GFP unfolding probabilities. (a) GFP unfolding probability plotted as a function of the ATP concentration. Note the x-axis is inverted. (b) GFP unfolding probability plotted as a function of ATPase rate (converted from ATP concentration). Note the x-axis is inverted.

2009; Moffitt et al., 2010), which is defined as the ratio of the squared mean of dwell times over the variance of the dwell times

\[
\begin{align*}
    n_{\text{min}} &= \frac{<\tau>^2}{<\tau^2> - <\tau>^2}.
\end{align*}
\]

(3.1)

It has been shown that this parameter provides a strict lower bound to the number of rate-limiting events during the dwell phase (Moffitt et al., 2009; Moffitt et al., 2010). We sought to understand how the number of rate-limiting events \(n_{\text{min}}\) changes with the ATP concentration. At saturating and near-\(K_m\) concentrations of ATP, we measured \(n_{\text{min}} = 2.1 \pm 0.4\) and \(2.0 \pm 0.6\) (SEM), respectively, suggesting that for both conditions there are at least two rate-limiting transitions in the dwell leading up to the burst phase. Similar values of \(n_{\text{min}} = 1.9 \pm 0.6\) were obtained at the intermediate ATP concentrations of 100 and 200 \(\mu\)M. Therefore, at least two processes—not associated with ATP binding to the titratable sites—control the duration of the dwell in the \([\text{ATP}]\) range between 35 \(\mu\)M (near \(K_m\)) and 5 mM (saturating).

Altogether, these results suggest a mechanism of translocation that is strikingly different from those of other motors. The well-characterized DNA packaging motor of the bacteriophage \(\phi 29\) exhibits a variable, ATP-concentration-dependent dwell time followed by a constant, ATP-concentration-independent burst size of 10 bp, which reflects a high degree of coordination among the ring subunits that must all load ATP before the motor can initiate translocation (Moffitt et al., 2009; Chistol et al., 2012). In sharp contrast, ClpXP exhibits an ATP-concentration-independent cycle time, during which a variable number of ATP molecules bind to the motor, resulting in a distribution of burst sizes.
3.2.4 ClpXP requires four highly coordinated power strokes for successful unfolding of GFP

Our findings that ClpX translocates with variable burst sizes and that the burst-size distribution is ATP-concentration dependent raises two important questions about the mechanism by which the motor successfully unfolds protein substrates. First, how are the power strokes of individual subunits coordinated during protein unfolding? And second, what is the kinetic competition between the bursts of motor translocation and the substrate’s resistance to unravel and its tendency to refold? To answer these questions, we quantified how variations in ATP concentration (and therefore in translocation burst sizes) affect the ability of the motor to unfold GFP. We found that reducing the ATP concentration decreased the probability of GFP unfolding by ClpXP nonlinearly from a maximum of 0.6 when [ATP] ≥ 500 µM to less than 0.1 when [ATP] ≤ 50 µM (Figure 3.7A).

The nonlinear relationship between GFP unfolding probability and ATP concentration suggests that GFP unfolding requires the coordinated and near-simultaneous ATP hydrolysis of multiple ClpX subunits in the ring. Supporting this hypothesis, the relationship between GFP unfolding probability and the ATPase rate of ClpXP also decreases nonlinearly with ATP concentration, in the same ATP concentration range where only the burst size distribution shifts to lower values during translocation (Figure 3.7B). In order to elucidate the coordination mechanism by which ClpXP successfully unfolds GFP, we characterized the intermediates observed during GFP unfolding. Unraveling of GFP from the C terminus proceeded via two transient intermediates with mean lifetimes of 45 ± 10 and 130 ± 15 ms (Figure 3.8A-D). By using the worm-like chain (WLC) model of polymer elasticity (Bustamante et al., 1994; Marko and Siggia, 1995), we estimated that the transition from the folded state “F” to the first intermediate “I” (F → I) has a contour length increase $\Delta L_{c,F \rightarrow I}$ of 8.3 ± 0.4 nm (SEM) (Figure 3.9A, blue) corresponding to the extraction of $\beta$ strand 11 ($\beta_{11}$) from the GFP barrel (Figure 3.10A, C). The second transition from “I” to the second intermediate “II” (I → II) has a $\Delta L_{c,I \rightarrow II}$ of 31.2 ± 0.8 nm (Figure 3.9A, B) and most likely corresponds to the unfolding of $\beta$ strands 10 through 7 (Figure 3.10B, C). The last transition from “II” to the unfolded state “U” (II → U) has a $\Delta L_{c,II \rightarrow U}$ of 42.2 ± 0.8 nm (Figure 3.9A, C) and reflects the unraveling of the remaining six $\beta$ strands, as described previously (Maillard et al., 2011) (see Section 3.4.7) The total contour length increase $\Delta L_{c,F \rightarrow U} = \Delta L_{c,F \rightarrow I} + \Delta L_{c,I \rightarrow II} + \Delta L_{c,II \rightarrow U}$ of 82.8 ± 3.2 nm (Figure 3.9A, D) is in agreement with the expected value for the complete unfolding of GFP (see Section 3.4.8).

At [ATP] ≤ 200 µM, we detected small unfolding and refolding events before ClpXP completely unraveled GFP (Figure 3.11A). The change in contour length during the reversible unfolding and refolding events is similar to $\Delta L_{c,F \rightarrow I}$, suggesting that these events most likely correspond to the extraction and quick refolding of $\beta_{11}$ (Figure 3.11B). Analysis of these reversible transitions revealed that $\beta_{11}$ snaps back into the GFP barrel with a mean refolding time constant $\sim$240 ms at forces between
Figure 3.8: GFP unfolds via two distinct intermediates. (a) The rips (sudden increase in bead extension) during GFP unfolding revealed three well-defined transitions, indicating the presence of two unfolding intermediates, labeled I and II, at saturating ATP concentrations. The second intermediate was observed in 90% of all traces (top right panel), while both intermediates were present in only 40% of them (top left panel). We rarely observed traces displaying only the first intermediate or no intermediates at all (10% of traces, bottom panels). (b)-(d) Distribution of lifetimes for the folded states, and the first and second intermediates during GFP unfolding by ClpXP. (b) shows the distribution of lifetimes of the folded state, which corresponds to the total GFP unfolding time. The distribution of lifetimes for the first and second intermediate (C and D) was well-fitted by a single-exponential function. The first intermediate has a lifetime of $45 \pm 10$ ms, whereas the second intermediate has a lifetime of $130 \pm 15$ ms. All fitted values are mean \( \pm \) SEM.

7 and 9 pN (Figure 3.11C). These results provide direct experimental evidence of the
Figure 3.9.: Estimating the contour length of GFP-unfolding intermediates using worm-like chain polymer theory. (a) Plot of force versus change in extension for the transition to the first intermediate $F \leftrightarrow I$ (blue), the second intermediate $I \leftrightarrow II$ (red) the unfolded state $II \leftrightarrow U$ (green), and the sum of all transitions $F \leftrightarrow U$ (black). (b)-(d) Histograms of contour-length change for the transitions.

molecular tug-of-war between the motor, attempting to unravel folded structures, and a substrate with a strong tendency to refold. Hence, protein-unfolding machines have to perform not only the thermodynamic function of mechanically destabilizing the native state, but also the kinetic task of quickly capturing the unstructured polypeptide before it can refold. A complete model for the successful ClpXP-mediated unfolding of GFP is shown in Figure 3.11D).

When $[ATP] \gg K_m$, ClpXP is able to move in bursts of 4 nm during a single translocation cycle. Such coordinated translocation is sufficient to trap most of the dislodged $\beta_{11}$ and prevent its refolding. These observations indicate that the efficient
unfolding of GFP by ClpXP requires not only a 4 nm burst, but also that this burst occur faster than the refolding time of $\beta_{11}$ ($<240$ ms). Using the distribution of burst size and dwell duration determined here, we estimated the probability of ClpXP taking a 4 nm burst in less than 240 ms to be on average 0.031 (0.018-0.046). This result indicates that ClpXP must pass through $\sim$33 translocation cycles before it can unfold GFP and trap the first intermediate for subsequent unfolding. Multiplying 33 translocation cycles by the mean duration of the dwell ($\sim$0.35 seconds) predicts a mean GFP unfolding time of $\sim$11.5 seconds, which agrees very well with the mean time constant obtained from the distribution of GFP unfolding times, $<\tau> = 11.8 \pm 0.9$ seconds (Figure 3.8D), as well as from previous bulk single-turnover GFP degradation measurements (Martin et al., 2008c).

At $[\text{ATP}] \sim K_m$, ClpXP moves in bursts of at most 3 nm, which may be sufficient to promote the extraction of $\beta_{11}$ from the GFP barrel (ClpX succeeds in carrying
Figure 3.11.: Intermediate ATP concentrations reveal failed \( \beta 11 \) extraction by ClpXP. (a) Trajectory at \([\text{ATP}] = 200 \ \mu \text{M}\) illustrating the ClpXP-induced unfolding and refolding of \( \beta 11 \). (b) and (c) Distributions of the change in contour length and the mean time constant for unfolding and refolding events of \( \beta 11 \). These events were observed in \( \sim 20\% \) of all traces at ATP concentrations near \( K_m \) and \( \sim 11\% \) of traces at intermediate \([\text{ATP}] = 100-200 \ \mu \text{M}\). At \([\text{ATP}] > 200 \ \mu \text{M}\), we hardly detected these events (less than \( \sim 1\% \) of all traces). (d) Mechanism of GFP unfolding by ClpXP at \([\text{ATP}] \gg K_m \).

out the thermodynamic task) but are too small to prevent the refolding of \( \beta 11 \) (ClpX fails to accomplish the kinetic task). Thus, under subsaturating ATP conditions, the kinetic competition between ClpXP attempts to translocate the unfolded region and the tendency of that region to refold greatly reduces the unfolding efficiency of the motor.

3.2.5 Hydrolysis-deficient ClpX rings translocate robustly but are marginalized in GFP unfolding

While titrations of ATP\( \gamma \)S and ATP both support a firing cycle that is largely independent of the filled state of the ring (after at least two subunits are filled with ATP), dissecting the particular order of nucleotide filling and hydrolysis activities remains inaccessible via titration studies. In order to probe these questions, mutant hexamers were constructed in which different numbers and orders of subunits were hydrolysis-deficient. The results discussed in this section are preliminary and require follow-up.
A previously characterized arginine-to-lysine mutation (R370K, referred to as an R subunit) in the sensor-2 motif (see Figure 1.2) of the ATP-binding pocket disables ClpX subunits from changing their conformation in response to ATP-binding, inhibiting nucleotide-dependent functions such as ClpP docking and substrate-ssrA engagement (Joshi et al., 2004). We used this mutation in a subset of subunits to observe what happens to ClpX-mediated protein degradation when a subset of subunits are inactivated and to also address whether the order of mutations matters (Figure 3.12).

Shown in Figure 3.13, the rate of translocation decreases as the number of consecutive R subunits increases. A maximum of three R subunits could be present within the ClpX ring, because substrate engagement and ClpP binding become compromised when more than half of the ring carries the sensor-2 mutation (Martin et al., 2005). The rate of translocation was determined by taking the total distance translocated and dividing this distance by the total amount of time taken by the motor, i.e., pauses were not removed.

As the number of R subunits increased, evidence for increased motor pausing became visually obvious (Figure 3.14). Upon removing these pauses (for pause-removal method, see Section 2.4.7), we observed that the pause-free velocity remained unchanged regardless of the number of mutant subunits within the ring (Figure 3.15). So, while the rate of translocation decreases roughly linearly with the number of mutant subunits, the pause-free velocity remains unchanged.

The decreased rate of translocation shown in Figure 3.13 can be explained solely by the increased frequency of pausing. The pause-density, or number of pauses per unit of translocation, increases from about 0.015 to about 0.05 nm\(^{-1}\) as a function of the number of consecutive R subunits present in the ring (Figure 3.16, orange bars).
Regardless of the pause density, however, the average duration of the pauses remains fixed at about 1.5 seconds (Figure 3.16, grey bars). Thus, a pause density of 0.015 nm\(^{-1}\) corresponds to one pause every \(\sim70\) nm translocated (\(\sim1\) pause per translocation through the first GFP and the following two I27\(^{CM}\) domains) whereas a pause density of 0.05 nm\(^{-1}\) corresponds to one pause every 20 nm (\(\sim4\) pauses per translocation through the same polypeptide region). The pausing behavior of the WWWRRR mutant thus adds about 5 extra seconds to the process of translocation which, on average, corresponds to about 80 nm of extension within the force range that these mutant studies were performed (7-9 pN). Since wild-type ClpX translocates at 8 nm/s, it usually takes 10 seconds to translocate through this region of the substrate. So the 5 extra seconds makes the rate of translocation about 40% slower, consistent with our observations shown in Figure 3.13.

Surprisingly, changing the order of sensor-2 ClpX subunits to RWWRWR, even with three ClpX subunits still carrying the sensor-2 mutation, rescues both the overall translocation rate as well as the pause-free velocity to values indistinguishable from wild-type ClpX (Figure 3.17A). As expected, the density of pauses recovers as well, consistent with the notion that the decreased translocation rate for the other mutants is solely caused by an increased frequency of pausing (Figure 3.17B).
Figure 3.14.: Representative traces showing pausing behavior observed during translo-
cation by ClpX sensor-2 mutants. (a) Representative trace between forces of 7-9 pN
for wild-type ClpX. (b) Representative trace between forces of 7-9 pN for ClpX con-
taining a single R subunit. Axes are identically scaled and dashed lines are parallel
to aid visual comparison.

Even with a single R subunit, the GFP-unfolding ability of ClpX becomes severely
compromised. Figure 3.18 summarizes the results of biochemical experiments mea-
suring the rate of GFP-I27\textsuperscript{CM}-ssrA degradation using the disappearance of GFP flu-
orescence as a readout for GFP unfolding/degradation (see Section 3.4.9 for method-
ology). In contrast to the translocation measurements, the mutant RWWRRRW does
Figure 3.15.: Pause-free velocity of translocation remains constant as a function of the number of consecutive R subunits. Error bars show standard error.

not recover GFP-unfolding capabilities similar to wild-type, indicating that the requirements for robust translocation are distinct from the requirements for protein unfolding.

3.3 Discussion

3.3.1 Intersubunit coordination determines the ClpX translocation mechanism

Titrations with the slowly hydrolyzable analog ATPγS as well as the analyses of burst sizes and dwell times have revealed several aspects of the intersubunit coordination in the ClpX hexamer. The pause-density dependence on the ATPγS concentration suggests that ClpX maintains operation even when one or two ATPγS molecules bind to the ring. At least three ATPγS molecules must bind to the ring in order to stall the motor. Based on previous biochemical and structural studies, at most four subunits in the hexamer can bind nucleotide (Hersch et al., 2005; Glynn et al., 2009; Stinson et al., 2013). Because binding of three analog molecules stalls the motor, the remaining fourth subunit may bind ATP but would still be unable
Figure 3.16.: The density of pauses during translocation increases as a function of the number of R subunits. Orange bars show the increase in pause density, or the number of pauses per unit if extension translocated, as a function of the number of R subunits within the hexamer. Grey bars show that, regardless of the number of R subunits, the average duration of the pauses remains constant. Error bars show standard error.

to drive translocation. This analysis suggests that the minimal operational unit for subunit translocation by ClpX involves at least two subunits.

That it takes binding of three analog molecules to stall translocation by the ClpX hexamer rules out models of strict intersubunit coordination described previously for other molecular motors (Lyubimov et al., 2011). For instance, it excludes models of concerted hydrolysis, where the subunit power strokes occur simultaneously after all subunits have been loaded with nucleotide. It also contradicts strictly sequential hydrolysis models, in which the power stroke of one subunit occurs only after the power stroke of its neighbor. If one of these models of intersubunit coordination were to apply to ClpX, a single ATP\textsubscript{γ}S binding event would be sufficient to stall the motor. Furthermore, our data are also inconsistent with stochastic or probabilistic hydrolysis models, where all subunits act independently from each other. In a stochastic scenario, all the active subunits would have to bind ATP\textsubscript{γ}S in order to stall translocation. Thus, we find that the ATPase cycles of individual ClpX subunits are neither strictly coordinated nor completely independent from each other. Instead, as will be discussed below, we find that two, three, or four subunits can coordinate their activity in each cycle, depending on the number of ATPs bound to the motor.

The analysis of burst sizes provides direct evidence for how many subunits participate in translocation during a single burst phase. The ClpX crystal structure reveals a distance of 1 nm between pore loops of adjacent subunits in different nucleotide states,
Figure 3.17.: Changing the arrangement of R subunits leads to full translocation recovery. (a) The translocation rate and the pause-free velocity are indistinguishable between wild-type ClpX hexamers and the asymmetric mutant RWWRWW. (b) The pausing characteristics between wild-type ClpX and the asymmetric RWWRWW mutant similar. Error bars show standard error.

suggesting a 1 nm power stroke per hydrolyzing subunit during substrate translocation (Glynn et al., 2009). Furthermore, dividing the maximum translocation velocity, $V_{max} = 8.5$ nm/s, by the rate of ATP hydrolysis, $k_{cat} = 8.3$ ATP/s, yields $1.02 \pm 0.03$ nm per hydrolyzed ATP. Thus, if one ATP is hydrolyzed per ClpX subunit, the fundamental translocation step size must be $\sim 1$ nm, consistent with previous single-molecule reports (Maillard et al., 2011; Aubin-Tam et al., 2011). Consequently, the variable burst sizes of 2, 3, or 4 nm observed at various ATP concentrations reflect the near-instantaneous, coordinated firing of two, three, or four subunits around the ring. The maximum burst size of 4 nm suggests that up to four subunits in the hexamer participate during a single translocation cycle. The smallest observed burst size as well as the minimum number of ATPγS molecules required to stall the motor allow us to propose that active translocation requires a minimum number of two coordinated subunits.

Surprisingly, our data reveal that bursts of 2 or 3 nm occur much more frequently than bursts of 4 nm, even at saturating concentrations of ATP, suggesting that the most relevant translocation cycle of the ClpX motor involves the coordinated hydrolysis and conformational change of two or three subunits. These results thus provide direct evidence for the operational flexibility of ClpX during polypeptide translocation. Previous bulk biochemical assays and crystal structures (Hersch et al., 2005; Glynn et al., 2009; Stinson et al., 2013) revealed hexamers bound to four nucleotides. These findings most likely reflect the state of the ring preceding a 4 nm burst, a state
Figure 3.18.: Even single R-subunit ClpX rings are severely compromised in GFP unfolding. Degradation activity is normalized to wild-type values. Despite the RWWRWW subunit exhibiting full recovery in translocation capability, it remains severerly compromised in it’s GFP unfolding ability.

that, in view of our results, is only one of a larger ensemble of nucleotide-bound configurations. Future structural studies in the presence of substrate may capture ClpX loaded with two or three ATP molecules, which according to our data correspond to conformational states that are more probable during polypeptide translocation. It may also be possible that, even if the ring binds four nucleotides, only two or three subunits fire near-simultaneously in a single burst of translocation, perhaps depending on the position at which hydrolysis or phosphate release originates. If the hydrolysis cycle progresses with preferred directionality, then it may be possible that a subset of nucleotide-bound subunits do not participate in a given translocation burst.

Our results show that the number of subunits participating in a single translocation cycle depends on the availability of ATP. When the ATP concentration was lowered from saturating to near $K_m$, we observed a redistribution of burst sizes from 3 and 4 nm to 2 and 3 nm. The predominant population of 2 nm bursts and the almost complete absence of 1 nm bursts indicates that, even at partly rate-limiting ATP concentrations ([ATP] $\sim$35 $\mu$M), at least two of the subunits must bind ATP in order to initiate a translocation cycle. Therefore, the $K_m$ of these two subunits that bind ATP first must be significantly lower than the average $K_m$ of 35 $\mu$M. The
lack of 4 nm bursts at ATP concentrations near 35 µM indicates that the remaining subunits have a higher $K_m$ than the first two high-affinity sites and thus contribute to translocation only at high ATP concentrations.

Two different classes of ATP binding sites with correspondingly different binding affinities have been previously identified in ClpX using traditional nucleotide competition assays (Hersch et al., 2005; Stinson et al., 2015). Furthermore, recent mutational studies of ClpX have shown strong evidence for a dynamic mechanism of subunit switching, whereby nucleotide binding can affect the affinities of the remaining subunits within the ring (Stinson et al., 2013; Stinson et al., 2015). Our data shed light onto how the coexistence of subunits with high and low ATP affinity affects the dynamics of the translocation cycle, allowing the burst phase to be initiated with either two ATPs (both binding to high-affinity subunits), three ATPs (two high- and one low-affinity subunits), or four ATPs (two high- and low-affinity subunits) bound to the motor.

A surprising finding is that, regardless of the number of subunits participating in the translocation cycle, the mean cycle time remains constant for ATP concentrations between $\sim K_m$ and saturating. The mean cycle time is expected to increase when ATP binding becomes rate limiting, and, to probe this behavior, we would have to monitor translocation in the presence of ATP concentrations at or below the $K_m$ for the high-affinity subunits ($K_m \ll 35$ µM). However, no translocation activity was observed for ATP concentrations below $\sim 29$ µM. Within the accessible ATP-concentration range, we found that the dwell duration is governed by at least two non-ATP-binding events ($n_{min} \sim 2$). The nonbinding events in the dwell could correspond to conformational changes within the ClpX ring that either (1) result from or (2) are completely independent of ATP binding.

### 3.3.2 An “internal clock” triggers polypeptide translocation

Our results require a model to rationalize the observed invariant dwell-time distribution and the variable burst-size distribution of ClpX as a function of ATP concentration. The dwell duration is largely determined by two slow, non-ATP binding events. By the time these slow transitions occur, the two high-affinity subunits are ATP bound and, with a probability that depends on the ATP concentration, one, two, or none of the low-affinity subunits are occupied with ATP as well. Due to the constrained ring geometry, only four of the six subunits can be ATP bound (Hwang and Lang, 2013), which provides an explanation for why we did not observe bursts of more than 4 nm even at saturating ATP concentrations. After the motor hydrolyzes all of the bound ATP molecules, Pi release and concomitant subunit power strokes occur near simultaneously around the ring, resulting in a burst size that is proportional to the number of hydrolyzed ATP molecules. The essence of this model is that the mean duration of the dwell phase is constant and set by an “internal signal” or “clock”, which may or may not follow ATP binding and could, for instance, correspond to the reaching of a strain threshold in the ring or the hydrolysis of the
Figure 3.19.: Schematic depiction of intersubunit coordination. The maximum number of firing subunits is different at saturating (blue box) versus limiting ATP concentrations (purple box). Model illustrates one possible scenario depicting sequential ATP binding. The subunits in gray correspond to those that do not bind/hydrolyze ATP. During the dwell phase, at least two ATPs can bind to the low-affinity ClpX subunits (T, blue outline), and additional ATPs can bind to the low-affinity ClpX subunits (T, green outline), depending on [ATP]. During the burst phase, the motor hydrolyzes all bound/committed ATPs, releases phosphate, and translocates the substrate by 2, 3, or 4 nm into the central pore.

First-bound ATP. In contrast to this constant average dwell phase, the burst size is variable and proportional to the number of subunits bound to ATP before the clock triggers the initiation of a translocation cycle around the ring. As a result, the ClpXP motor operates at a constant frequency (fixed “rpm”) and a variable burst size (different “gears”). Although our current model, depicted in 3.19, suggests a spatial and temporal order of ATP-docking events, the catalytic mutant ClpX studies described in Section 3.2.5 strongly favor a probabilistic model of ATP-hydrolysis.
ClpX hexamers containing one, two, or three subunits with R370K mutations in the sensor-2 motif of the ATP-binding pocket all showed unaffected polypeptide translocation velocities. While consecutive R subunits did exhibit an increased frequency of pausing, changing the order of mutations could result in a full recovery of wild-type translocation characteristics, even though half of the ring was unable to hydrolyse ATP. So, while the motor does indeed exhibit strong plasticity in the order and number of subunits that can load ATP, there does seem to be a preferred order of how subunits coordinate their individual ATP binding and hydrolysis cycles. Even a single R subunit present in the ClpX hexamer severely inhibited the ring’s ability to robustly unfold GFP. Surprisingly, the RWWRRW mutant ring, which was indistinguishable from wild-type during translocation, was still unable to robustly unfold GFP. These results effectively demonstrate that the subunit-coordination requirements for polypeptide translocation significantly differ from the subunit-coordination requirements for effective protein unfolding.

3.3.3 Biological relevance of ClpX translocation mechanism

The model proposed here provides a framework to understand how ClpXP successfully unfolds stable protein substrates. The ability of the motor to bind four ATP molecules allows it to translocate in large bursts and thus destabilize and rapidly trap partially unfolded intermediates. For instance, to successfully unfold GFP, ClpX subunits must near-simultaneously take a 4 nm burst, which results in the extraction and translocation of β11 from GFP before this strand can refold back onto the β barrel of GFP. At saturating concentrations of ATP, the time required for ClpX-induced GFP unfolding is determined by the time that passes before the motor makes a 4 nm burst. In contrast, ClpX rarely unfolds GFP at ATP concentrations near $K_m$ because hydrolysis under these conditions is always triggered before four ATP molecules can bind to the motor. In support of our finding, previous biochemical studies have shown a nonlinear decrease in GFP degradation with decreasing ATPase rate (Martin et al., 2008c), whereas a linear correlation emerges for circular permutants of GFP that do not form a stable unfolding intermediate upon extraction of their C-terminal β strand (Nager et al., 2011). Our results reveal that the decreased probability of the motor taking a 4 nm burst (from ~20% at saturating [ATP] to ~0% at [ATP] near $K_m$) is responsible for the observed nonlinear decrease of GFP unfolding probability with the ATPase rate.

Based on bulk biochemical studies, it has recently been suggested that the ClpX motor may operate by a partially probabilistic mechanism during unfolding and translocation (Stinson et al., 2013). Our results show that ClpX translocates substrates using a highly coordinated mechanism in which, regardless of the number of translocating subunits, the average dwell duration is constant and followed by a variable burst size that reflects the firing of a different number of subunits in rapid succession. Although there is a stochastic element during the loading of ATP, we
observed a high degree of intersubunit coordination during the rapid burst phase of translocation.

Such a mechanism of a constant cycle time (resulting in a variable burst size) allows flexibility for ClpX to successfully overcome unique chemical and mechanical obstacles during polypeptide translocation. A constant cycle time governed by an internal clock provides the motor with a fail-safe mechanism to drive translocation even when some subunits are not loaded with nucleotide. This characteristic will allow the motor to prevent substrate disengagement or prolonged periods of stalling during unfolding attempts when a subset of subunits is unable to maintain grip on the polypeptide.

Our results thus provide an archetype of molecular-motor operation that differs significantly from those described previously for other ring-shaped motors. For instance, the translocation mechanism of the φ29 DNA packaging motor requires that periodic contacts be made by the motor every 10 bp along the helical pitch of double-stranded DNA. In order to achieve a fixed burst size of 10 bp, the motor must wait during the dwell until all subunits are loaded with ATP. This mode of operation may have been optimized to keep the motor in register with the symmetry of its DNA substrate and provides a fascinating contrast to our results on ClpX. Because a polypeptide track is aperiodic, ClpX cannot rely on contacting a regularly repeating motif, and its variable burst size (different “gears”) and constant cycle time (constant “rpm”) may have arisen as a flexible mechanism that optimizes its efficiency for robust kinetic trapping of unfolding intermediates. The distinction between these mechanisms offers clear evidence for the evolutionary constraints imposed by the motor’s substrates to favor certain mechanisms of operation. The mechanisms described here thus provide important insights into the operating principles of ATP-dependent proteases and may have critical implications for the understanding of other ring-shaped ATPases of the AAA+ and RecA families in general.

3.4 Materials and Methods

3.4.1 Measurements of ATP-γS-hydrolysis rate

The ATP-γS-hydrolysis rate of ClpXP was measured using thin-layer chromatography of 35S-labeled ATP-γS (Burton et al., 2003). Hydrolysis was measured at room temperature by mixing 0.3 μM single-chain ClpX hexamer and 1.5 μM ClpP with variable concentrations of ATP-γS (5-800 μM) and trace amounts (70 nM) of 35S-labeled ATP-γS (Perkin-Elmer Inc.) with or without ssrA-titin(10 μM). One-microliter aliquots from each 20 μL reaction were removed at different times and immediately quenched with 2.5 μL of stop buffer (50 mM Tris pH 7.6, 100 mM EDTA, 20 mM ATP-γS, and 20 mM Na3PO4). Each quenched sample was spotted onto a plastic-backed PEI-cellulose sheet (J.T. Baker Inc.) and chromatographed at room temperature in 1.5 M formic acid and 0.4 M LiCl. After the sheet was dried, radioactivity in each spot was quantified using a Molecular Dynamics Phosphoimager. In the control
experiment with only \(^{35}\)S-labeled ATP\(\gamma\)S, we observed that the single radioactive spot, corresponding to the un-hydrolyzed \(^{35}\)S-ATP\(\gamma\)S, hardly migrated from its loading position. In the presence of ClpXP, we observed a second, faster-migrating radioactive spot whose intensity increased in a time-dependent fashion (Figure 3.1). This spot corresponds to PO\(^{35}\)S that is released by ClpXP upon ATP\(\gamma^{35}\)S hydrolysis, and its absence in samples without ClpX indicates that spontaneous hydrolysis of ATP\(\gamma\)S is negligible under our experimental conditions.

### 3.4.2 Extracting pauses from ATP\(\gamma\)S traces

The general pause-detection scheme was the same as described in Section 2.4.7. ATP\(\gamma\)S-induced pauses were extracted by removing ATP-only dwells using a double-exponential fit with one time constant fixed to the mean dwell duration measured at saturating [ATP].

### 3.4.3 Motivation for fitting ATP\(\gamma\)S-induced pauses to the Hill equation

The reaction pathway for an ATP\(\gamma\)S-induced pause can be written as

\[
M_a + n\text{ATP}\gamma\text{S} \leftrightarrow M_p,
\]

where \(M_a\) is the actively translocating state of the motor, \(M_p\) is the ATP\(\gamma\)S-induced paused state of the motor bound to the analog, and \(n\) is the number of ATP\(\gamma\)S molecules required to induce a pause. The apparent dissociation constant \(K_d\) of \(n\)ATP\(\gamma\)S molecules binding to the motor can be expressed as

\[
K_d = \frac{P(M_a)[\text{ATP}\gamma\text{S}]^n}{P(M_p)}
\]

where \(P(M_a)\) is the probability of the motor being in an actively translocating state and \(P(M_p)\) is the probability of the motor being in an ATP\(\gamma\)S-induced paused state. The pause density \(PD([\text{ATP}\gamma\text{S}])\), or number of pauses per unit length of translocated polypeptide, should be directly proportional to \(P(M_p)\). Because intrinsic pauses in the absence of ATP\(\gamma\)S are infrequent - on average one pause per 50 nm of translocated polypeptide (Maillard et al., 2011) - the approximation is appropriate

\[
P(M_p) + P(M_a) = 1.
\]

Using Equation 3.4 in Equation 3.3 to eliminate \(P(M_a)\) results in

\[
PD([\text{ATP}\gamma\text{S}]) \propto P(M_p) = \frac{[\text{ATP}\gamma\text{S}]^n}{K_d + [\text{ATP}\gamma\text{S}]^n}.
\]

Hence, \(PD([\text{ATP}\gamma\text{S}])\) takes the form of the Hill equation with \(n\) reflecting the number of ATP\(\gamma\)S molecules that must bind to the motor in order to induce a pause.
3.4.4 Calculating the fraction of ATPγS molecules bound to the hexamer

For a ClpX motor subunit M present in solution with a mixture of ATP and ATPγS (denoted as A and B, respectively), the pathway for the subunit binding either nucleotide can be written as

\[ MB \leftrightarrow B + M + A \leftrightarrow MA. \] (3.6)

The total number of motor subunits in the ensemble is constant and equals

\[ [M]_T = [M] + [MA] + [MB]. \] (3.7)

Thus, the fraction of A and B molecules bound to M with their corresponding dissociation constants rewritten in terms of \( f_A \) and \( f_B \) as

\[ f_A = \frac{[MA]}{[M]_T} ; K_d^A = \frac{[M][A]}{[MA]} = \frac{(1 - f_A - f_B)[A]}{f_A} \] (3.8)

and

\[ f_B = \frac{[MB]}{[M]_T} ; K_d^B = \frac{[M][B]}{[MB]} = \frac{(1 - f_A - f_B)[B]}{f_B} \] (3.9)

Equations 3.8 and 3.9 constitute a system of equations that can be used to obtain expressions for \( f_A \) and \( f_B \) in terms of \([A], [B]\), and the dissociation constants \( K_d^A \) and \( K_d^B \).


and


From our biochemical studies of ATP and ATPγS hydrolysis in the presence of substrate, \( K_m \) can be used as an upper-bound estimate for the \( K_d \) because \( k_{cat} \) has a small contribution. From the hydrolysis experiments described above (Figure 3.1), the \( K_m \) values were determined to be 57 µM and 29 µM for ATP and ATPγS, respectively. Therefore, we can use the \( f_A \) to estimate the probability of \( n \) ATPγS molecules bound to the available six active sites in the hexamer with the binomial distribution

\[ P(n) = \binom{6}{n} p^n q^{6-n}, \] (3.12)

where \( p = f_A \) and \( q = 1 - f_A \) which includes both populations bound to ATP and not bound to any nucleotide.
3.4.5 Step-fitting methodology

Steps were fit using methodology previously described (Moffitt et al., 2009;Maillard et al., 2011; Chistol et al., 2012). Regions of traces with high signal-to-noise (S/N) exhibited step-like patterns in translocation regions (Figure 3.20). Stepping correlations were confirmed by visualizing the pair-correlation function of data points within the selected region, which shows the distribution of pair distances between data points (Figure 3.21, top right). Confirmed regions of stepping were fit using a sliding window t-test to detect the changepoints that occur during the burst phase of translocation (Figure 3.21, bottom right).

3.4.6 Spatial/temporal resolution of the instrument

The lifetime distributions of each intermediate during ClpX-mediated GFP unfolding (Figure 3.8) follow a single-exponential decay function. To calculate the probability of observing an intermediate with duration \( \tau \) we must first transform this function into its corresponding probability density function, which is expressed as

\[
f(x) = ae^{-at},
\]

where \( a \geq 0 \) and corresponds to the inverse of a time constant in seconds, and \( x = [0, \infty) \). The probability of observing a GFP-unfolding intermediate whose lifetime is between 0 and \( \tau \) can be calculated by obtaining the area under the curve given by the following expression
Figure 3.21.: Step-fitting methodology. Step-like pattern is confirmed using pair-correlation function (top right). The steps are fit using a sliding-window t-test (bottom right). See Section 3.6 for details.

\[ P(0 \leq x \leq \tau) = \int_{0}^{\tau} ae^{-ax} dx. \]  

(3.14)

Using the time constants from the exponential fit of the distributions in Figure 3.8, we can use this expression to test if the apparent absence of the first or the second intermediate in some traces is because (1) their lifetime may be shorter than the detection limit of the instrument, or (2) GFP unfolding may follow multiple pathways where the first or second intermediate may not be obligatory transitions in this process.

To address the first scenario, we established the spatial/temporal detection limits of our instrument. We calculated the fluctuations of the noise at different forces and different bandwidths through the analysis of the power spectrum of a trace where ClpXP translocation has been stalled with 100% ATPγS. Because we analyzed the GFP-unfolding intermediates at 300 Hz (i.e. 300 data points per second), we found
that fluctuations of the noise at this bandwidth are approximately 3 nm. With this value, we calculated the signal-to-noise ratio (SNR), which corresponds to the size of the signal divided by the size of the fluctuations (standard deviation) of the noise. We included this SNR in the following expression,

$$\Delta l \geq \sqrt{\frac{2k_B T N \gamma}{\langle \tau \rangle \kappa_{tether}}} \text{SNR},$$

which predicts the minimal duration of a step that is required to be observed given its size. Here, $\Delta l$ corresponds to the mean size of the step, $k_B$ is the Boltzmann constant, $T$ is the temperature, $N$ is the number of uncorrelated measurements per average dwell, $\gamma$ is the drag coefficient of the beads, $\langle \tau \rangle$ is the mean duration of the dwell, $\kappa_{tether}$, and SNR is the signal-to-noise ratio that we have calculated (Moffitt et al., 2008). Using this expression for a burst size of 8.3 nm (corresponding to the mean extension of the first rip), we found that its corresponding mean dwell duration (first intermediate duration) should be at least 30 ms to be observed at 300 Hz. With this threshold, we calculated the theoretical probability of not observing an intermediate due to limitations of the instrument by calculating the probability $P(0 \leq x \leq 30)$ using the expression in Equation 3.14. Finally, we compared this theoretical probability with the experimental probability of not observing the first or second intermediate in our traces. In other words, based on their lifetimes, we estimated the probability of missing the first and second intermediates in the unfolding trajectories to be 0.52(0.44, 0.5) and 0.15(0.13, 0.19), respectively, where each range represents 95% confidence on the prediction. The predicted numbers were in good agreement with the fraction of single-molecule traces in which we did not observe either intermediate (0.65 ± 0.11 and 0.15 ± 0.03 for the first and second intermediate, respectively). Thus, the simplest conclusion is that the four different unfolding scenarios described in Figure 3.8A result from a single ClpXP-mediated GFP unfolding pathway that has two obligatory intermediates rather than a branched unfolding pathway. We conclude that the subset of GFP-unfolding intermediates not observed in our traces where due to instrument-sampling limitations.

### 3.4.7 Identifying the structures of ClpXP-mediated GFP-unfolding intermediates

The crystal structure of GFP (PDB ID: 1GFL) was used to map the observed rip size of each unfolding transition observed in our traces to the corresponding extraction of GFP structural elements using a previously established methodology (Dietz and Rief, 2004; Ormö et al., 1996; Maillard et al., 2011). The observed change in contour length $\Delta L_c$ could be written as

$$\Delta L_c = m_i L_{aa} - \left( X_{m_i+1:N}^F - X_{m_{i-1}+1:N}^F \right).$$

(3.16)
The first term $L_c = m_i L_{aa}$ describes the actual contour length of each unfolding transition, where $m_i$ is the number of residues that unfold during the transition to the $i$th intermediate, and $L_{aa}$ is the contour length of one amino acid (0.365 nm/aa). The second term comprises the shift in the distance between the last structured residue at the N-terminus (residue position 1) and the last structured residue at the C-terminus as the transition occurs. Both distances $X_F^{m_i+1:N}$ and $X_F^{m_{i-1}+1:N}$ are determined from the reported crystal structure coordinates.

Equation 3.16 provides an iterative means to determine which structural elements of GFP were extracted during each intermediate transition (see Figure 3.10A and B). First, the force at which each unfolding transition occurred was plotted as a function of its corresponding change in extension in nanometers. Using the worm-like chain model of polymer elasticity (Bustamante et al., 1994), a force-extension curve was generated to calculate the $L_c$ of an unfolding transition occurring at each amino acid position from Equation 3.16 (iterate m). Selecting the contour length $L_c$ that minimized the root-mean-square deviation (RMSD) from the WLC fit allowed us to establish the optimal location of the unfolding segment for each intermediate (see structural mapping in Figure 3.10C).

This approach provided an unambiguous identification of the first intermediate, namely, a state with beta-strand 11 ($\beta_{11}$) extracted from the GFP barrel. For the second intermediate, however, the RMSD versus amino acid position yielded two minima with similar RMSD values: one solution corresponded to the extraction of $\beta_{10-7}$ and a second solution to the extraction of $\beta_{10-6}$. A simple line of reasoning allowed for distinguishing between these two possibilities. When GFP unfolding occurred in the absence of the first small intermediate (Figure 3.8A, right top and bottom), the RMSD versus amino acid position had a single minimum corresponding to the extrac-
tion of β11-7. Importantly, the distribution of intermediate lifetimes was identical to the distribution of second-intermediate lifetimes when the first intermediate was observed (Figure 3.22), suggesting also similar structural stability of the unfolded intermediate. Thus, the most parsimonious explanation is that the second unfolding transition both in the presence and absence of the first one corresponds to the same structural intermediate, with β11-7 unraveled from the GFP barrel.

3.4.8 Prediction of GFP unfolding time based on burst size and dwell time distributions

Using the distribution of burst size and dwell duration (Figure 3.6), we estimated the probability of ClpXP taking a 4 nm burst in less than 240 ms (mean refolding time of β11 from Figure 3.11) to be on average 0.031(0.018, 0.046), where the range represents 95% confidence. This result indicates that ClpXP must pass through on average ~33 translocation cycles before it can unfold GFP and trap the first intermediate for subsequent unfolding. Multiplying 33 translocation cycles by the mean duration of the dwell (~0.35 seconds) predicts a mean GFP unfolding time of 11.5(7, 21) seconds, where the range represents 95% confidence. This value agrees remarkably well with the mean time constant obtained from the distribution of GFP unfolding times, \( \langle \tau \rangle = 11.8 \pm 0.9 \) seconds (Figure 3.8B).

3.4.9 GFP-fluorescence degradation assay

ClpXP complex was reconstituted from purified single-chain ClpX and ClpP in degradation buffer (25 mM HEPES pH 7.6, 20 mM MgCl₂, 300 mM KCl, and 0.5 mM ETA) in the presence of ATP-regeneration system. Upon addition of ybbr-GFP-(I27CM)₄-ssrA substrate, degradation activities were monitored by the loss of GFP fluorescence (excitation 488 nm; emission 509 nm) using a QuantaMaster spectrofluorimeter (PTI). Reaction conditions where as follows: 0.3 µM ClpX, 1.5 µM ClpP, and 20 µM substrate.
4. Concluding Remarks and Future Directions

This thesis has revealed that ClpXP translocates polypeptides in a unique manner. The ClpX ring behaves as a variable-displacement motor, capable of triggering bursts of translocation with different numbers and configurations of ATP-bound subunits. The duration of the rate-limiting step(s) that actually trigger phosphate-release and thus translocation are independent of the number of filled subunits (after two subunits have bound ATP). Mutational studies also suggest that this duration is largely independent of the spatial configuration of filled subunits, but more careful analysis will be required to definitively address this key question. Although we have yet to identify the transitions that set the duration of ClpX’s dwell phase, the rich structural information available provides testable candidates for future studies. For example, the dwell-determining transitions may follow the binding of ATP and involve setting up at least one subunit for catalysis, perhaps by correctly positioning the arginine finger from the neighboring subunit (Kainov et al., 2008). The rigid interfaces between subunits could then facilitate the coordinated and near-instantaneous firing of the remaining nucleotide-bound subunits, regardless of their actual configuration.

These results thus provide an archetype of molecular-motor function that differs significantly from those described previously in other ring-shaped motors. These differences may reflect evolutionary constraints that tailor different motors for their specific functions in the cell. For instance, the translocation mechanism of the φ29 DNA packaging ATPase requires that periodic contacts be made every 10 bp along the helical pitch of double-stranded DNA as the motor pushes its viral DNA into its attached viral capsid. In order to achieve this fixed burst size, the motor must wait for all subunits to load ATP before the burst phase of translocation can be initiated (Moffitt et al., 2009). This mode of operation may have been “optimized” to keep the motor in register with the symmetry of its DNA substrate. In contrast, ClpXP encounters no such symmetry. Because a polypeptide track is aperiodic, ClpX cannot rely on contacting a regularly repeating motif. Furthermore, ClpX has to also be capable of unfolding stably-folded proteins. Its variable burst size and constant cycle time may have arisen as a flexible mechanism that optimizes its efficiency at performing both its polypeptide translocation and protein unfolding tasks. The clear distinction between these motor’s purposes and the details of their functionality offers fascinating evidence for how evolutionary constraints can bias molecular function.

Although this work was focused on understanding the detailed mechanism of ClpXP-mediated protein degradation, future studies may confirm that homologous ring-shaped motors function in a similar manner. Cryo-EM reconstructions of the 26S
proteasome (a eukaryotic AAA+ protease that shares a similar overall architecture and function with ClpXP), point to a similar ring-architecture, including the formation of rigid-body interfaces between subunits as well as a staircase arrangement of pore-1 loops, during the actively-translocating state of the machine. Rigid bodies between neighboring subunits, as observed in ClpX crystal structures, were found to be tilted to different extents within the plane of the unfoldase ring, leading to a continuous spiral-staircase arrangement of the six pore-1 loops. The fixed position of each distinct subunit within this spiral may originate from the heterohexameric nature of the ring or the asymmetry imposed by surrounding ubiquitin-interacting modules and the proteasomal peptidase itself. Substrate translocation could thus in theory proceed through only local pore-1 loop motions with subunits otherwise statically positioned in the spiral arrangement. Alternatively and more likely, the spiral staircase observed in the EM reconstructions may represent a long-lived state of an ATPase ring, whose hydrolysis cycles rely on a biphasic mechanism involving a long-lived, stationary dwell phase followed by a burst-phase where subunits rapidly undergo conformational motions that generate a “power-stroke” of translocation.

Both ClpX and the φ29 DNA packaging motor have been observed to spend more than 90% of their time in the dwell phase and less than 10% of their time undergoing conformational changes during the burst phase of translocation. A potentially similar behavior of the 26S proteasome as well as other ring-shaped motors could explain why their ATPase subunits appear to be arranged in a static spiral staircase in crystal structures (where the static state is the energetically-favored state) as well as cryo-EM reconstructions (where thousands of particles are averaged during reconstruction). Notably, the pseudo two-fold symmetric structure of ClpX shows a different organization of pore-1 loops, in which each half ring contains a separate three-step staircase (Glynn et al., 2009). However, the ClpX structure was solved in the absence of substrate, and it is conceivable that this motor adopts a continuous spiral staircase, similar to the proteasome unfoldase, as soon as substrate is engaged and being translocated. Of course, it cannot be ruled out that ClpX and the proteasome use different mechanisms for translocation, and future studies will be necessary to definitively distinguish between these possibilities. Besides the cryo-EM reconstruction of the 26S proteasome, the structures of only three other substrate-bound ring translocases have been resolved to date: the AAA+ E1 helicase (Enemark and Joshua-Tor, 2006) as well as the structurally related RecA-type Rho and DnaB helicases (Thomsen and Berger, 2009; Itsathitphaisarn et al., 2012). Interestingly, similar to the proteasome, each of these structures shows a continuous spiral arrangement of subunits around the ring (Nyquist and Martin, 2014), further suggesting that unfoldases like ClpX may also adopt such a staircase arrangement upon substrate engagement in the central pore.

Remarkable progress has been made in understanding the polypeptide translocation and protein unfolding mechanism of ClpXP as well as other ATP-dependent proteases in recent years. The emerging models have required the convergence of interdisciplinary fields spanning state-of-the-art approaches in structural biology, bulk
biochemistry, and single-molecule force spectroscopy. Of course, fundamental questions still persist, such as figuring out the identity of the transitions that control the dwell time during substrate translocation and mapping out the detailed spatiotemporal ordering of ATP binding, hydrolysis, and product release events within the ATPase hexamers. Addressing these questions using ClpX as the most advanced model system will likely provide important new insight into the operating principles applicable to other ATP-dependent proteases and ring-shaped motors serving different functions in the cell. However, it will be critical to determine whether our current understanding of the ClpX mechanism can indeed be generalized. Recent hints from related motors of the AAA+ and RecA families, such as the 26S proteasome and DNA helicases, suggest considerable similarities in basic function, and a mechanistic convergence of the ATP-dependent motor field is an exciting prospect.
A. Fitting steps in passive-mode optical tweezers experiments

A.1 Introduction

The high spatial and temporal resolution afforded by single-molecule optical tweezers allows recording of the kinetics of individual molecular motors in real time. Of course, the accuracy of resulting kinetic models is only as reliable as the data analysis techniques, and unfortunately, data analysis is particularly challenging for single-molecule trajectories where the signal-to-noise is low. When the data are noisy, many kinetic models may be reasonable and the ultimate goal is to select the simplest one, adding complexity only as necessary.

For analysis of ClpXP stepping patterns discussed in Chapter 3 we used a methodology described in Section 3.6 (Block and Svoboda, 1995; Carter and Cross, 2005; Moffitt et al., 2009; Chistol et al., 2012). This method has the advantage of transparency, because only those steps that are clearly observed are included within the overall analysis scheme, i.e. there is no doubt that the signal carries a step-like pattern. However, methods such as these can hold several severe drawbacks as well. For instance, there’s a high potential for selection bias towards larger step sizes and longer dwell times, because those are the transitions easiest to visually detect. There’s also the serious drawback that analyses can not easily be reproduced, since the method is fundamentally manual in nature. Furthermore, visual recognition of steps is time-consuming and therefore, as the number of experiments increases, the amount “human-time” required for analysis becomes untenable. For example, a quick back-of-the-envelope calculation can illustrate why stepping analysis was difficult to perform with the catalytic mutants described in Section 3.2.5. Consider analyzing the step-pattern of 10 different catalytic mutants. In order to carefully find steps within each trace using the method described in Section 3.6, visual inspection and analysis requires \(\sim 30\) minutes of “human time”. Furthermore, in order to obtain appropriate statistics, between 250 and 500 steps must be detected. If, for instance, the [ATP]-dependence of each mutant would be interesting to study, that would add about 10 different conditions per mutant. Carrying this through, our prediction for the time required to perform such an analysis manually would have a lower bound of \(\sim 12,500\) hours. Combined with the inherently slow process of collecting optical tweezers data, these time commitments are simply too large to accomplish the work.

A reasonable goal, then, would be to design automated algorithms for detecting steps in noisy time series data; a goal which has spawned a major area of research...
within the field of single-molecule biophysics (Carter et al., 2008). In general, these approaches tend towards a search for a series of statistically significant jumps within the data. We call these series of discrete steps, attributed to the molecular motor, a “model”. The model statistics can subsequently be used to construct a kinetic model to describe the underlying mechanism of the motor. While in principle the only input of such an algorithm should be the trajectory’s noise statistics, there are many different ways these statistics are handled in practice, from simple thresholding and pre-filtering techniques to more elaborate and theoretically rooted methodologies (Kalafut and Visscher, 2008; Little and Jones, 2013; Kerssemakers et al., 2006).

In this appendix, we will consider a case such as data obtained from passive-mode optical tweezers experiments, where the noise statistics themselves vary in time. We will build off of a previously developed algorithm (Kalafut and Visscher, 2008), which works quite effectively for situations in which the noise statistics are identically, independently, and normally distributed (Figure A.1). In cases where the noise statistics do not reflect Kalafut’s assumptions (i.e. they are non-stationary or correlated), Kalafut’s algorithm fails in a potentially disastrous manner (Figure A.2), to the extent that wrong conclusions can be drawn regarding the mechanism of the motor. While passive-mode experiments generate data that are difficult to analyze, they have the benefit of not requiring the extra instrumentation and extra error sources that constant-force experiments introduce. I choose to include this section as an appendix rather than a chapter because I am not yet satisfied by its performance in real-world situations. There are many useful concepts, however, and this subject represents a major interest area of mine. Ultimately, I think a much more powerful step-fitting algorithm could be created by simultaneously handling both non-stationary noise and correlated noise.

The algorithm developed in this appendix does not represent an end-all-be-all solution to the step-fitting problem faced by many single-molecule biophysicists. In fact, other “Achilles heel” aspects of noise-statistics exist in optical tweezers experiments that will not be considered here, such as the correlated data points described above (Arunajadai and Cheng, 2013). Instead, I hope that the general approach proves valuable to the curious reader and that aspects of the algorithm may someday become incorporated within a generalized solution to the problem of noisy step detection.

A.2 Algorithm

Our goal here is to adapt the Schwartz Information Criterion (SIC) (Schwartz et al., 1978; Kalafut and Visscher, 2008) to situations where we aim to detect steps in single-molecule trajectories with non-stationary noise statistics. The SIC is a criterion for model selection which discriminates between possible models, given the data. The optimal model is chosen as the one that minimizes the SIC. Thus, the optimal model is both (1) the simplest and (2) provides a good fit to the data. Both of these
component 1 tempers component 2’s tendency to increase model complexity in order to improve fitting.

A.2.1 Derivation

Here, we will derive the Schwartz Information Criterion directly from Bayes’ theorem. For clarity we will use the language of single-molecule optical tweezers data. Consider the following definitions. $M_k$ is the $k^{th}$ possible single-molecule stepping
Figure A.2.: Fitting steps using SIC to a trace with identically, but not independently distributed Gaussian noise fails impressively. A similar stepping trajectory to that shown in Figure A.1 was generated with S/N=1, except that the overlaid Gaussian noise is identical but not independently distributed (i.e. data points are correlated). Again, the magnitude of the signal was taken to be the step size and the magnitude of the noise was taken to be the standard deviation of the overlaid noise. Correlations in the noise were generated according to the methodology described in Section A.5.1 and exponentially decay with $A = \tau/400$. The simulated data is shown in blue and the “true” model in black. The model detected using Kalafut’s algorithm is shown in red. Because of the correlated noise, the likelihood function does not adequately reflect the noise statistics and the SIC fails dramatically at detecting the correct steps. This is a realistic problem encountered during the analysis of optical tweezers data, as the data are often correlated.

Bayes’ theorem states that

$$p(M_k, \theta^k|Y)p(Y) = p(Y|M_k, \theta^k)p(M_k, \theta^k),$$  \hspace{1cm} (A.1)
where $p(M_k, \theta^k | Y)$ is the posterior probability density of the model given the data, $p(Y)$ is the probability density of the data, $p(Y | M_k, \theta^k)$ is the probability density of the data given the model (known as the likelihood $L(\theta^k | Y)$ of the model), and $p(M_k, \theta^k)$ is the prior probability density of the model parameters given a particular model.

Rearranging Equation A.1, we obtain

$$p(M_k, \theta^k | Y) = p(Y)^{-1} L(\theta^k | Y) p(\theta^k, M_k), \quad (A.2)$$

where, again, the likelihood function is defined as $L(\theta^k | Y) = p(Y | M_k, \theta^k)$. The probability of the model is therefore obtained by integrating both sides of Equation A.2 over the model parameters $\theta^k$

$$P(M_k | Y) = p(Y)^{-1} \int d\theta^k L(\theta^k | Y) p(\theta^k, M_k). \quad (A.3)$$

The optimal model is the one which maximizes $P(M_k | Y)$, or equivalently which minimizes $-2 \log P(M_k | Y)$, which is the SIC (Schwartz et al., 1978). The SIC is often expressed in the limit of large $n$ and where the data points are sufficiently independent such that the likelihood of the entire data set is simply the product of independent likelihoods over each data point

$$L(\theta^k | Y) \sim \exp \left( -nV(\theta^k) \right). \quad (A.4)$$

In other words, the likelihood that a model produced the observed data is roughly the product of the likelihoods that the model produced each individual data point. Series expansion of the likelihood about its maximum, $\hat{\theta}^k$ gives

$$L(\theta^k | Y) = \exp \left( -nV(\hat{\theta}^k) - \frac{n}{2!}(\theta^k - \hat{\theta}^k) \cdot V^{(2)}(\hat{\theta}^k) \cdot (\theta^k - \hat{\theta}^k) + \cdots \right), \quad (A.5)$$

where $V^{(2)}(\hat{\theta}^k)$ is a second derivative of $V$ evaluated at $\hat{\theta}^k$. Placing the second-order expansion shown in Equation A.5 into Equation A.3 and performing the multidimensional Gaussian integral, we find

$$-2 \log (P(M_k | Y)) = -2 \log \left( \exp \left( -nV(\hat{\theta}^k) \right) \ast \left( \frac{2\pi}{n V^{(2)}(\hat{\theta}^k)} \right)^{D_k/2} \right)^{D_k/2} - 2 \log p(\hat{\theta}^k, M_k) + \log p(Y). \quad (A.6)$$

The SIC is simplified form of Equation A.6 where two types of terms are dropped: (1) terms that do not explicitly vary from model to model and (2) terms which are assumed to scale slower than $O(n^0)$. The SIC is therefore

$$SIC = -2 \log L(\theta^k | Y) + D_k \log n, \quad (A.7)$$
where $\mathcal{L}(\theta^k|Y) = \exp \left( -nV_n(\theta^k) \right)$. Thus for large $n$, the functional form of the prior $p(\theta^k, M_k)$ is irrelevant and dropped. However, when $n$ is small enough, the prior can shift the location of the posterior’s maximum, potentially leading to the selection of a different, more optimal stepping model.

For non-stationary, passive-mode optical tweezers trajectories we can imagine slicing the trace into segments, each with approximately equal noise statistics. So, while we can consider the noise within an individual segment approximately stationary, the noise statistics over multiple segments could vary drastically. This concept will become the heart of our algorithm.

Consider a single segment with $n$ data points having $d_k$ edges (or steps) detected in the data at point $j_1, \cdots, j_k$. We define the first step to be at the first data point ($j_0 = 0$) and the last step to be at the last data point ($j_{k+1} = n+1$). With this notation, the $i^{th}$ step contains $n_i$ data points and runs from point $j_i$ through point $j_{i+1} - 1$.

For Gaussian identical and independently distributed noise, the likelihood function is parametrized in terms of $\theta^k = (\mu, \sigma_{j_1\cdots j_k})$ with $\mu = \{\mu_i\}$ the collection of means and $\sigma_{j_1\cdots j_k}$ the standard deviation over all steps within this segment (Kalafut and Visscher, 2008). The likelihood function is

$$
\mathcal{L}(\theta^k|Y) = \prod_{i=0}^{d_k} \prod_{l=j_i}^{j_{i+1} - 1} \frac{1}{\sigma_{j_i\cdots j_k} \sqrt{2\pi}} \cdot \exp \left( -\frac{(y_l - \mu_i)^2}{2\sigma_{j_i\cdots j_k}^2} \right). 
$$

We bias the standard deviation of the noise $\sigma_{j_1\cdots j_k}$ (not the mean location of the detected steps) using the following prior (Lee, 2012),

$$
p(\theta^k, M_k) = p(\sigma_{j_1\cdots j_k}^2 \equiv \phi, M_k) = N(S_0, \nu) \phi^{-\nu/2-1} \exp \left( -\frac{S_0}{2\phi} \right) \tag{A.9}
$$

where the parameters $\nu$ and $S_0$ are the prior sample size and prior standard deviation, respectively, and $N$ is a normalization constant. The parameters $\nu$ and $S_0$ can be determined experimentally, from real-time control traces held at different force-equivalents for each segment,

$$
E(\phi) = \frac{S_0}{\nu - 2} \tag{A.10}
$$

$$
V(\phi) = \frac{2S_0^2}{(\nu - 2)^2(\nu - 4)}. \tag{A.11}
$$

$E(\phi)$ represents the expected (or average) value of the control signal’s variance and $V(\phi)$ represents the variance of the control signal’s variance over many different measurements.
Inserting the likelihood and prior (Equations A.8 and A.9) into the expression for $P(M_k|Y)$ (Equation A.3) and performing the integration we recover

$$P(M_k|Y) = p(\{Y\})^{-1} \cdot N(S_0, \nu) \cdot \frac{(2\pi)^{-(n-d_k)/2}}{n_1^{1/2} \cdots n_{d_k}^{1/2}} \cdot \frac{1}{2} \cdot \frac{(S + S_0)^{-(n+\nu+d_k-1)/2}}{2} \cdot \frac{n + \nu - d_k - 3}{2}! \quad (A.12)$$

where $S = n_1 \hat{\sigma}^2 + \cdots + n_{d_k} \hat{\sigma}_{d_k}^2 \equiv n \hat{\sigma}^2$ and where $\hat{\sigma}_i^2$ is the variance of the data points about the $i^{th}$ step

$$\hat{\sigma}_i^2 \equiv \frac{1}{n_i} \sum_{l=j_i}^{j_{i+1}-1} y_l^2 - \frac{1}{n_i^2} \left( \sum_{l=j_i}^{j_{i+1}-1} y_l \right)^2. \quad (A.13)$$

To clarify, $n_i$ is the number of data points in the $i^{th}$ step whereas $n$ is the number of total data points in the considered segment with approximately stationary noise. So, if there are $d_k$ steps, $n = n_1 + \cdots + n_{d_k}$.

The model that maximizes $P(M_k|Y)$, or equivalently, minimizes $-2 \log P(M_k|Y)$ is the optimal model. When $n$ is sufficiently large we can expand the factorial using Stirling’s approximation. By dropping terms which are not model dependent we obtain an expression we shall call the SICP, or the Schwartz Information Criterion, with prior

$$SICP = -d_k (\log 2\pi + 1) + \log n_1 + \cdots + \log n_{d_k}$$

$$+ (n + \nu - d_k - 1) \log (n \hat{\sigma}^2 + S_0) - (n + \nu - d_k - 3) \log (n + \nu - d_k - 3)$$

$$- \log (n + \nu - d_k - 3) + O(n^{-1}) \quad (A.14)$$

Indeed, under Kalafut’s assumptions of constant prior (i.e. $S_0$ and $\nu = 0$) as well as that all steps are roughly the same size ($n_i \sim n/d_k$ for all $i$) and that $d_k$ scales slower than $\log n$, we recover the familiar

$$SIC = n \log \hat{\sigma}^2 + d_k \log n + O(n^{-1}). \quad (A.15)$$

None of these conditions hold true for our ClpXP experiments (and for many experiments performed by other single-molecule biophysicists), and we shall observe a performance improvement when we use the SICP vs. the SIC for data that simulates passive-mode experiments. For ClpXP, we have shown that steps of different sizes occur based upon the filled state of the ring. Also, in general, the number of detected steps should increase linearly with the number of data points (i.e. a single-molecule trajectory that’s twice as long as another should, on average, have twice as many steps). Also, of course, passive-mode experiments carry non-stationary noise statistics.
Figure A.3.: Depiction of SICP algorithm. Data figure taken from (Kalafut and Visscher, 2008) with the addition of color “slices” to depict how a trace with non-stationary noise would be segmented. The top shows the best single-step fit, the middle the best two-step fit, and the bottom the final, optimal fit.

A.2.2 Computational scheme for model selection

We follow Kalafut’s general scheme for detecting steps, locating them one at a time using all of the data (Kalafut and Visscher, 2008). Instead of calculating the SIC statistic under the assumption of overall stationary noise, however, we segment the data into small slices, where the noise is approximately stationary (Figure A.3). The overall model-selection criterion is simply the sum of the individual SICP values for each individual slice. For example, if the data are sliced into 5 segments, the overall SICP would be

\[(SICP)_T = SICP_1 + \cdots + SICP_5.\]  \hspace{1cm} (A.16)

The algorithm is quite simple, and mimics the general schema of Kalafut’s as follows:
1. Calculate the \((SICP)_{T}\) under the assumption there are no steps.

2. For each possible location of a single step in the data, calculate the \((SICP)_{T}\) (using the entire dataset, of course). Accept the location for which the \((SICP)_{T}\) is minimized as the best one-step fit.

3. If the \((SICP)_{T}\) for the one-step hypothesis is smaller than the \((SICP)_{T}\) for the no-step hypothesis, continue. Otherwise terminate and conclude that no steps exist.

4. Holding the position of the first step fixed, calculate the \((SICP)_{T}\) for each possible location of a second step in the data. Accept the location for which the \((SICP)_{T}\) is minimized as the best two-step fit.

5. If the \((SICP)_{T}\) for the two-step hypothesis is smaller than the \((SICP)_{T}\) for the no-step hypothesis, continue. Otherwise terminate and conclude that there is only a single step.

6. Continue in this fashion, adding one step at a time until the \((SICP)_{T}\) is no longer minimized by an additional step.

Figure A.3 illustrates the functionality of the \(SICP\) algorithm, with the actual trace taken directly from (Kalafut and Visscher, 2008) and the colored slices representing the segments of data where the noise-statistics are assumed to be stationary.

A.3 Results

The SICP is a much more effective step-fitting algorithm for situations in which the noise statistics are non-stationary in time. We randomly generated 200 traces with a step size of 3 units and an average dwell duration of 300 data points. The overlaid noise was Gaussian and linearly increased in S/N after each step such that the beginning of the trace had S/N=1 and the end of the trace had S/N=3. Each trace contained 800 data points per unit of time. Figure A.4 shows a representative trace (blue) with the “true” model shown in black, the SICP fit shown in red, and the SIC fit shown in yellow. The SIC overfits significantly in the beginning of the trace (low S/N) as expected because the SIC assumes stationary noise. The SIC fails to strike the correct balance between complexity and simplicity. In contrast, the SICP method handles the non-stationary noise explicitly and provides a much more stable fit. As shown in Figure A.5, the SICP method is much more stable across the entire range of the traces, which provides a basis for demonstrating its validity over the SIC method.

Figure A.5A shows the step size histograms resulting from the SIC algorithm (top) and the SICP algorithm (bottom) for the first third of the 200 simulated traces. The simulated traces have only a single forward step (Figure A.5C), so the backwards steps observed in Figure A.5A (top) are the artifacts resulting from overfitting. In contrast,
Figure A.4.: SICP effectively detects steps in traces with non-stationary noise. (a) A stepping trajectory with non-stationary noise was generated. The noise level was decreased by a constant amount after each step such that the S/N change from 1 at the beginning of the trace to 3 after 15 steps (the end). The magnitude of the signal was taken to be the step size and the magnitude of the noise was taken to be the standard deviation of the Gaussian noise. The “true” fit is shown in black. The fit obtained using the SICP algorithm is shown as a dashed red line, whereas the yellow line shows the fit obtained by Kalafut’s algorithm. (b) Zoomed-in view showing the beginning of the trace. (c) Zoomed-in view showing the end of the trace.

The SICP fits (Figure A.5A, bottom) are quite stable and much more representative of the true model. Figure A.5B shows the step size histograms resulting from the SIC algorithm (top) and the SICP algorithm (bottom) for the last third of the 200 simulated traces. Here, the SIC and the SICP fits look essentially identical and are both representative of the true model.

The extent of overfitting observed in Figure A.5A is worrisome because it could lead to wrong conclusions regarding the mechanism of a particular molecular motor. Despite the histogram showing an approximately bipartite distribution of forward and backward steps, the simulated traces only have forward steps. This artifact serves as an important reminder that conclusions drawn from single-molecule analysis are strongly limited by the analysis techniques employed.
Figure A.5.: Histograms of step sizes detected in traces with non-stationary noise. Column (a) shows the SIC results, column (b) the SICP results, and column (c) the “true” model. 200 different traces were generated identically to that shown in Figure A.4. The top histograms show step size distributions measured from the first third of the traces in time (lowest S/N regions). The bottom histograms show step size distributions measured from the last third of the traces in time (highest S/N regions).

Next, we can test the behavior of the SICP and the SIC in situations where the number of data points are small, even when the noise statistics are stationary. We randomly generated 200 traces with a step size of 3 units and an average dwell duration of 300 data points. The overlaid noise was Gaussian and stationary. However, each trace only had 200 total data points, violating the SIC’s (i.e. Kalafut’s) assumption of large \( n \). We find that the SIC significantly overfits, finding structure in the data that does not exist (Figure A.6A, yellow). In contrast, due to the prior that introduces our expection for the noise statistics, the SICP finds step models that much more closely approximate the truth (Figure A.6A, red). Leaving the low \( n \) limit, we find that the SIC and the SICP converge to identical solutions, as expected ((Figure A.6B).

A.4 Discussion

The SICP method developed above is an approximate tool for finding steps in traces with non-stationary noise statistics. In passive-mode experiments the stan-
Figure A.6.: SICP effectively detects steps in traces with a small number of data points. (a) Traces with small numbers of data points (200 data points per trace). Left: Representative stepping trajectory showing fits obtained using the SICP algorithm (red), the SIC algorithm (yellow), and the “true” model (black). Middle and Right: 200 different traces were generated identically. The middle histogram shows the step size distribution measured using the SIC algorithm. The histogram at the right shows the step size distribution measured using the SICP algorithm. (b) Same as in (a) but with large number of data points (10000 per unit time).

Standard deviation for different regimes of applied-external force can be biased according to different priors constructed from control experiments. Although the method introduces two parameters, they are measured ones. We demonstrate that misuse of step-fitting algorithms, for instance, when their underlying assumptions are violated by the data, can lead to qualitatively wrong conclusions regarding the mechanism of molecular motors.
A Bayesian approach to step fitting can be used to bias the search for steps according to our belief for how the standard deviation should change within a passive-mode trajectory. We can also construct our prior to include our belief of how the autocorrelation function of the data points should look. A combination of these beliefs could result in a very powerful step-fitting algorithm and this is definitely a direction I would like to move in. While we focused our attention on passive-mode optical tweezers experiments, this type of approach is valuable in other fields, such as fluorophore counting, where the noise varies as a function of the number of fluorophore (and equivalently, the number of photons detected).

A.5 Methods

A.5.1 Simulating exponentially correlated Gaussian noise

We assume that the autocorrelation of the data follows

$$\langle x(t)x(t') \rangle - \langle x(t)^2 \rangle \propto e^{-A|t-t'|}, \quad (A.17)$$

where $A$ is the decay constant (i.e. the number of data points before the intensity of the correlation decays be a factor of $e^{-1}$). Assuming that correlations only accumulate within dwells and that memory is “reset” once a new state is reached, then

$$x(t) = \sigma_d \sum_{i=0}^{t} e^{-A|i-t|} X_i + \mu_d, \quad (A.18)$$

where $\sigma_d$ is the standard deviation of the noise, $\mu_d$ is the position of the dwell, $t$ is the number of preceding data points in the dwell, and $X$ is a vector of iid Gaussian random variables of length equal to the number of data points in the dwell and with mean $= 0$ and variance $= 1$. 
REFERENCES


